A GPT no-go theorem for the classicality of the gravitational field

The theoretical minimum

Andrea Di Biagio*
Institute for Quantum Optics and Quantum Information (IQOQI) Vienna, Austrian Academy of Sciences, Boltzmanngasse 3, A-1090 Vienna, Austria and Basic Research Community for Physics e.V., Mariannenstraße 89, Leipzig, Germany
(Dated: December 12, 2023)

The aim of this lecture is to cover the minimal theoretical machinery that goes into the proof of the following no-go theorem:

Theorem 1 (No-Go Theorem)**.** Let A and B be two GPT systems interacting with the gravitational field G. Then thee following three statements are incompatible:

- (GIE) A and B can become *entangled*,
- (Subsystem Locality) G *mediates* the interaction between A and B,
- (Classicality) G is a *classical* system.

This theorem is a generalisation due to Thomas Galley, Flaminia Giacomini, and John Selby [1], of the arguments made in [2] and [3] in favour of the claims that the observation of Gravity Induced Entanglement (GIE) is evidence for the non-classical nature of gravity. It is a proof within Generalised Probabilistic Theories (GPT), a *theory of theories*, a framework commonly used in the study of the foundations of quantum mechanics. By proving a result such as the one above within a framework, one can rule out entire families of theories.

I. OPERATIONAL ASPECTS OF QUANTUM THEORY

While the interpretation of quantum theory is still a hotly debated issue [4], there is one thing that everyone agrees to: quantum theory is remarkably successful in predicting the outcomes of measurements [5] and so while people disagree on whether we can think of the quantum state $|\psi\rangle$ as representing some objective state of affairs, we all agree on how to extract the *empirical* content of the quantum state: *probabilities* and, more generally, the *expectation values* of various mathematical operators corresponding to physically measurable quantities. Besides one's metaphysical preferences, thinking of the quantum formalism as first and foremost a tool for computing probabilities can be illuminating, as many features of Hilbert space quantum mechanics become clearer. Entire books can be (and, indeed, *have been*) written about this topic. Here we will focus on some basic things. The introductory chapters of [6] are a great resource for all this.

A. Density operators and the partial trace

For the purpose of computing probabilities, the correct quantum state to assign to a system is not a normalised vector in Hilbert space, but a *density operator* (aka density matrix):

Definition I.1. *A density operator ρ is Hermitian operator with all positive eigenvalues such that* $\operatorname{tr}\rho = 1$.

There are two ways to motivate this. First, suppose that believe that normalised states are the correct way to assign a quantum state to your system, but you are not quite sure what state to assign. This could be because of noise or because you are a bad experimentalist, or any other reason. Say you believe that your system is in the quantum state $|\psi_i\rangle$ with probability p_i, then when you compute the expectation value of some operator $\hat{O}$, you will have to first conditionalise on your ignorance of the "actual" state:

$$\langle\hat{O}\rangle = \sum_i p_i \langle\psi_i|\hat{O}|\psi_i\rangle = \operatorname{tr}\rho\hat{O}, \tag{1}$$

* andrea.dibiagio@oeaw.ac.at

where we have defined the density operator

$$\rho = \sum_i p_i \left|\psi_i\right\rangle\!\left\langle\psi_i\right|, \tag{2}$$

and the trace operation

$$\operatorname{tr} \hat{O} := \sum_i \langle i|\hat{O}|i\rangle, \tag{3}$$

where $|i\rangle$ is any orthonormal basis (doesn't matter which) for the Hilbert space.

This way, it looks like you only need the density operator formulation when you have some ignorance of the "true" state, and you can get along without them otherwise. However, when dealing with composite systems, the situation is basically unavoidable. Indeed, say that you are considering two systems A and B, but for some reason you can only access A. For example, B is some really hard or expensive to measure system. Given that the Hilbert space associated with the composite system is the tensor product of their respective space $\mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B}$, this means you can only measure operators of the form $\hat{O}_\mathsf{A} \otimes \hat{\mathbb{I}}_\mathsf{B}$. Assume that the systems are in some state $|\psi\rangle$. Any state in $\mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B}$ can be written in the form

$$|\psi\rangle = \sum_m r_m |a_m\rangle |b_m\rangle, \tag{4}$$

for some orthonormal bases (ONB) $\{|a_m\rangle\}$ and $\{|b_m\rangle\}$ and for some non-negative real numbers r_m whose squares sum to 1. This is known as the Schmidt decomposition, and it is an extremely handy result. We then have that

$$\langle\psi|\hat{O}_A|\psi\rangle = \sum_{mn} r_m r_n \langle a_m|\hat{O}_\mathsf{A}|a_n\rangle \langle b_m|b_n\rangle = \sum_n r_n^2 \langle a_n|\hat{O}_\mathsf{A}|a_n\rangle = \operatorname{tr} \rho_\mathsf{A} \hat{O}_\mathsf{A}, \tag{5}$$

where in the last step we defined the density operator

$$\rho_\mathsf{A} = \sum_n r_n^2 \left|a_n\right\rangle\!\left\langle a_n\right|. \tag{6}$$

We notice that B dropped out, but it seems that we need to compute the Schmidt form of our state to compute properties of A. However there is a fast way to get ρ_A, and this is by using the *partial trace operation*:

$$\operatorname{tr}_\mathsf{B} \rho := \sum_m \langle b_m|\rho|b_m\rangle, \tag{7}$$

where $\{|b_m\rangle\}$ is a ONB for $\mathcal{H}_\mathsf{B}$, (again, it doesn't matter which one). Then you can easily show that

$$\rho_\mathsf{A} = \operatorname{tr}_\mathsf{B} \left|\psi\right\rangle\!\left\langle\psi\right|. \tag{8}$$

Therefore, whenever you are interested on observables on only part of a system, you can use the reduced density matrix $\rho_\mathsf{A} := \operatorname{tr}_\mathsf{B} \rho$.

Since, in general, apparata are noisy, and systems are always subsystems of other systems, it is more appropriate and operational to use density operators as states for quantum systems.

Now it is a good time to recall the distinction between pure and mixed states. A pure quantum state is one that can be written in the form $\rho = |\psi\rangle\!\langle\psi|$, all the others are mixed. This is another way to see that the density operator state space is more general than the vector in Hilbert space. If you are handed a big density matrix, it is not necessarily clear wether it can be put into this form. There is however a very simple test of purity:

Proposition I.1. A quantum state ρ is pure if and only if $\operatorname{tr} \rho^2 = 1$.

The state space of a quantum system is a *convex set*, meaning that if you have two have two density operators ρ_1 and ρ_2, then a *convex* (or *stochastic*) *combination* $p\rho_1 + (1-p)\rho_2$, where $p \in [0,1]$ is also a valid state. This seemingly innocuous mathematical property is central in GPTs. The physical interpretation is that if you can prepare two states, you can have a machine that prepares either state at random. The convex structure also offers an equivalent characterisation of purity:

Proposition I.2. A state is pure if and only if it cannot be written as a nontrivial convex combination of other states.

Exercise I.1. Check that the result of taking the trace or the partial trace is independent of the basis used in the summation.

Exercise I.2. Read up on the Schmidt decomposition.

Exercise I.3. Prove Proposition I.1.

Exercise I.4. Prove Proposition I.2.

B. State space of a qubit

Let us consider as an example the state space for qubits. The space of density operators for a qubit is the set of 2×2 positive matrices. There is a neat way to visualise it. The 2×2 identity matrix

$$\mathbb{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{9}$$

and the Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{10}$$

form a basis for self-adjoint 2×2 matrices. If $\vec{r}$ is a 3-component vector such that $|\vec{r}| \leq 1$, then the matrix

$$\rho = \frac{1}{2}\mathbb{I} + \frac{1}{2}\vec{r} \cdot \vec{\sigma}, \tag{11}$$

is a unit trace, positive operator. Thus, the space of density operators of a qubit is isomorphic to a 2 dimensional ball, the *Bloch ball*. The boundary ($|\vec{r}| = 1$) consists of the pure states, and the inside the mixed states.

Exercise I.5. Compute $\operatorname{tr} \rho$ and $\operatorname{tr} \rho^2$ for ρ in (11). Compute its eigenvalues and eigenvectors. Compute the expectation value of spin measurements.

Exercise I.6. Show that the state space of the qubit is convex, by checking that if ρ_1 and ρ_2 are of the form (11), then so is a convex combination of them.

Exercise I.7. Show that a mixed state of a qubit can be written as a convex combination of pure states in many different ways. You may consider, without loss of generality, the state $\frac{1}{2}\mathbb{I} + \frac{1}{2}z\sigma_3$.

C. Sending states to states: instruments and CPTP maps

Since the density operator is a natural generalisation the state vector in the context of classical uncertainty and multipartite systems, and if we are interested in arbitrary manipulations of a given system, it is good to study possible maps between density operators on their own terms.

First, let us allow that the input and output Hilbert spaces $\mathcal{H}_{\text{in}}$ and $\mathcal{H}_{\text{out}}$ are different. This way, we can represent operations such as adjoining a system, or ignoring parts of the system. Since quantum experiments yield different outcomes probabilistically, we represent an evolution by a *set* of maps $\mathcal{E} = \{E_i\}$ from positive operators on $\mathcal{H}_{\text{in}}$ to positive operators on $\mathcal{H}_{\text{out}}$, where i labels one of the mutually exclusive outcomes. We ask that $\operatorname{tr} E_i\rho$ is the probability $p(i|\rho, \mathcal{E})$ of the outcome i to happen and thus that

$$0 \leq \operatorname{tr} E_i\rho \leq 1 \quad \text{and} \quad \sum_i \operatorname{tr} E_i\rho = 1, \tag{12}$$

for all density operators ρ. Next, we ask each of the E_i to be *convex linear*, meaning that

$$E_i\big(p\rho + (1-p)\sigma\big) = pE_i\rho + (1-p)E_i\sigma \tag{13}$$

for any two states ρ and σ and probability p. This is so that the probabilities given by $\operatorname{tr} E_i\rho$ behave consistently with stochastic mixtures. Finally, we want the state after applying E_i to still be a positive operator. This requires that E_i is a *completely positive* map. *Positive* means that $E_i\rho$ is a positive operator on $\mathcal{H}_{\text{out}}$ whenever ρ is a positive operator on $\mathcal{H}_{\text{in}}$. *Completely positive* means that, for arbitrary $\mathcal{H}_{\text{A}}$, $(I_{\mathcal{H}_{\text{A}}} \otimes E_i)$ is a positive map from the operators on $\mathcal{H}_{\text{A}} \otimes \mathcal{H}_{\text{in}}$ to those on $\mathcal{H}_{\text{A}} \otimes \mathcal{H}_{\text{out}}$, where $I_{\mathcal{H}_{\text{A}}}$ is the identity map on operators of $\mathcal{H}_{\text{A}}$. The requirement of complete positivity ensures that the applying $\mathcal{E}$ to a subsystem always yields a well-defined state of the combined system.

In sum, each of the maps E_i is a *trace non-increasing, completely positive* (CP) map. If the outcome i attains, then the state is updated to

$$\rho \longmapsto \frac{E_i\rho}{\operatorname{tr} E_i\rho}, \tag{14}$$

which is again a density operator. The set $\mathcal{E} = \{E_i\}$ is called an *instrument*, or *quantum operation*. We can also define a map induced by $\mathcal{E}$ itself: if we do not know or care about the result, we weight each of the outcomes above by their probability $p(i|\rho, \mathcal{E}) = \operatorname{tr} E_i\rho$ and then we have

$$\mathcal{E}\rho = \sum_i p(i|\rho, \mathcal{E}) \frac{E_i\rho}{\operatorname{tr} E_i\rho} = \sum_i E_i\rho. \tag{15}$$

Any evolution with a single outcome, like $\mathcal{E}$, or an instrument with a single outcome, is known as a *quantum channel*, and is represented by a *completely positive, trace preserving* (CPTP) map. The formalism of positive operators and completely positive maps is the most general way to formulate the evolution of quantum systems.

Let us connect this rather abstract formalism to unitary evolution. If a system undergoes a unitary evolution U, then its density operator changes as

$$\rho \longmapsto U[\rho] = U\rho U^\dagger. \tag{16}$$

The map U is then known as a *unitary channel*. Suppose instead that the system S under consideration interacts with another system E initially in some σ, so that the evolution of the combined system is a unitary channel, and then the second system is ignored. The resulting evolution for the state ρ of the initial system is

$$\rho \longmapsto \operatorname{tr}_{\mathsf{E}} \left[U[\rho \otimes \sigma]\right]. \tag{17}$$

This is a CPTP map. Indeed, *any* CPTP map can be represented this way, with σ a pure state, a result known as Stinespring dilation [**?**]. Say that the ancillary system instead is subjected to a projective measurement represented by the projectors $\{P_k\}$, then each map

$$E_k : \rho \longmapsto \left[(I \otimes P_k)U[\rho \otimes \sigma]\right]. \tag{18}$$

is a trace non-increasing CP map, and together they form an operation. Thus every instrument can be understood as the system interacting with an ancilla, and then the ancilla being measured. This result is known as Ozawa dilation [**?**].

Exercise I.8. Check that the unitary channel in (16) is trace preserving.

Exercise I.9. Check that the E_k in (18) indeed form an instrument.

D. Entanglement

The definition of entanglement is as straightforward as it is unilluminating: a state is *entangled* if it is not *separable*. A vector in Hilbert space $|\psi\rangle \in \mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}}$ is said to be separable if it can be written as

$$|\psi\rangle = |\psi_A\rangle \, |\psi_B\rangle \,, \tag{19}$$

while a density operator on $\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}}$ is separable if it can be written as a mixture of product states:

$$\rho = \sum_i p_i \, \rho_{\mathsf{A}}^{(i)} \otimes \rho_{\mathsf{B}}^{(i)}. \tag{20}$$

Great, so what? The first thing to note is that it is not trivial at all to tell if any density operator is an entangled one just by staring at it. Fortunately, for pure states, there is a super fast way to tell:

Proposition I.3. A pure state ρ of a composite system $\mathsf{A} \otimes \mathsf{B}$ is entangled if and only if the reduced density operator $\mathrm{tr}_{\mathsf{A}}\, \rho$ is mixed.

This makes precise one weird aspect of entanglement: if we think of a pure state as a state of maximal knowledge, then it is possible to have maximal knowledge about a composite system but have partial—and, in the limit, minimal—information about its parts.

For *mixed states*, determining whether a state is separable is much harder and finding necessary and sufficient conditions for entanglement in general systems is still work in progress [7]. However there are many *sufficient* conditions for entanglement. One such test is the *positive partial transpose* (PPT) criterion.

Definition I.2. *The* partial transpose ρ^Γ *of a bipartite state* ρ *is obtained by transposing on only one of the subspaces:*

$$\langle im|\rho^\Gamma|jn\rangle = \langle in|\rho|jm\rangle\,. \tag{21}$$

Proposition I.4 (PPT criterion)**.** If ρ^Γ is a positive operator, then ρ is separable [8].

For states of $\mathbb{C}^2 \otimes \mathbb{C}^2$ and $\mathbb{C}^2 \otimes \mathbb{C}^3$, this condition is both necessary and sufficient [9]. To experimentally use the PPT criterion, one needs to obtain the full density matrix, which requires many measurements to determine. Often it is cheaper to measure an *entanglement witness* [9, 10].

Definition I.3. *An* entanglement witness W *is an observable such that*

$$\mathrm{tr}\, W\sigma \geq 0 \tag{22}$$

for all separable states σ *but*

$$\mathrm{tr}\, W\rho_0 < 0 \tag{23}$$

for at least one entangled state ρ_0*.*

It follows that any state such that $\mathrm{tr}\, W\rho > 0$ is an entangled state. Thus, measuring a negative expectation value for W in the lab implies[1] that the state generated is an entangled state.

Exercise I.10. Prove Proposition I.3. Hint: use the Schmidt decomposition.

E. Local operations and classical communication cannot create entanglement in quantum theory

A well-known result in quantum information theory is that two spatially separated agents, each acting on their own quantum system, cannot create entanglement between their systems, even if they coordinate their actions by transmission of classical information [7].

The argument is quite simple. A *local operation* is represented by a map $\mathcal{E} \otimes \mathcal{F}$, where $\mathcal{E}$ and $\mathcal{F}$ are channels. The channels could represent an instrument where the outcome is ignored. This is because the two parties are not allowed to communicate, so they cannot coordinate their behaviour. Allowing one transmission of classical information from one party to the other, allows the choice of one operation conditional on the result of the other. This situation is one round of LOCC (local operations and classical communications) and is represented by a map

$$\sum_i E_i \otimes \mathcal{F}^{(i)}, \tag{24}$$

where $\{E_i\}$ is an instrument and each $\mathcal{F}^{(i)}$ is a channel. The $\mathcal{F}^{(i)}$s have to be deterministic channels as the second party is not allowed to send the result of their operation to the first party in this scenario. The map representing a

---

[1] The intuition behind this is pretty simple. We recall the map $(A, B) \longmapsto \mathrm{tr}\, AB$is a scalar product on the space of Hermitian operators, making the latter a normed real vector space. The set of S separable density operators is a convex subset of this space. It follows that for any point ρ_0 not in S, there is a hyperplane separating S from ρ_0. Then every point on the same side of the hyperplane as ρ_0 is automatically not in S.

round of LOCC where the second party sends information to the first is defined similarly. If the state of the system is initially separable, it will still be separable after one round of LOCC, since

$$\sum_i \left(E_i \otimes \mathcal{F}^{(i)}\right)(\rho \otimes \sigma) = \sum_i E_i\rho \otimes \mathcal{F}^{(i)}\sigma = \sum_i p_i\, \rho_i \otimes \sigma_i. \tag{25}$$

Thus an initially separable state will remain separable after any number[2] of LOCC rounds.

This result was used in [2] to argue that the detection of GME would prove that gravity is a quantum system. However, as pointed out by [3], one might distrust an argument based completely on quantum theory, as we do not know that gravity obeys either quantum or classical laws, and might follow a new set of laws. They argue that one ideally needs a similar argument in a more general framework. This is achieved in [11] and [1]. We will focus on the latter in the coming sections II and III.

II. GENERALISED PROBABILISTIC THEORIES

We now broaden our horizons to consider a framework that generalises the operational formulation of quantum theory. The main idea behind operationalism is that any scientific theory should—at the very least—provide probabilistic predictions about laboratory procedures. One might argue that a good theory scientific theory should also provide an explanation, or a picture of what is going on in nature, and tell us something about the world outside our laboratories, but a theory cannot be a good scientific theory if it can't tell us what we should expect to see in a given experiment.

This idea has led to the development of a few related mathematical frameworks, most notably Generalised Probabilistic Theories (GPTs) [12], Operational Probabilistic Theories (OPTs) [13], and Process Theories [14], and, more recently, Constructor Theory [15]. These are frameworks in which the predictive content of different physical theories can be formulated and compared. They are of great use in studying the information-processing capabilities of different theories, in a similar way that, say, linear algebra is useful in studying the formal properties of physical theories as different as quantum mechanics and fluid dynamics: knowing that your theory is a OPT with such and such properties allows you to immediately derive a host of results. For example, as we will shortly see, in every GPT it is impossible to send a signal without exchanging a system.

We will focus on the minimal aspects of GPTs required to understand the theorem. For more thorough introduction to GPTs, see these notes by Markus Müller [16], or the seminal paper by Johnathan Barret [12].

A. States, transformations and effects

Mathematically, a GPT system S consists of a convex set of states Σ_S, a convex set of transformations T_S and a convex set of effects E_S, which map states to real numbers. In general there will be a vector space V_S associated with S, such that states are vectors, effects are dual vectors, and transformations are linear maps $V_\mathsf{S} \to V_\mathsf{S}$. States, effects and transformations are called *processes*.

Intuitively, this generalises the structure of the operational formulation of quantum theory we saw above, where the state space is convex, evolutions map states to states, and there are other maps that allow you to compute probabilities. Quantum theory is a paradigmatic example of a GPT theory, with states as density matrices, CPTP maps as transformations and applying CP maps and taking followed by tracing as effects. One simple way GPTs differ from each other is in the shape of the convex set of states and effects.

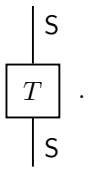
Operationally, the states Σ_S correspond to *equivalence classes of preparations* of the system (those that yield the same probabilities), the transformations T_S correspond manipulations (including observations) that can be done to the system without destroying it or losing it and effects E_S represent manipulations (including observations) after which the system is then destroyed or simply ignored.

GPTs come with an expressive diagrammatic calculus in which systems are represented by (labelled) wires

(26)

[2] Note the actual result is a bit stronger than this, see [7] for details.

and processes are represented by boxes with dangling wires

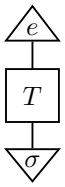


. (27)

In particular, states have no input wires and effects have no output wires:

. (28)

A diagram with no free dangling wires such as



(29)

corresponds to a number (a state in the trivial GPT system).

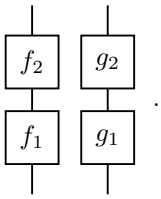
The diagrammatic calculus becomes particularly useful when one starts considering different systems and how they interact. Two GPT systems can be composed in parallel by using the tensor product structure of the associated spaces. The resulting system is then represented by wires side by side

. (30)

The multi-system diagrammatic calculus provides two related advantages over traditional "1D" formulas. First, a formal advantage, since it makes redundant a number of equations relating parallel and sequential composition. For example, the property

$$(f_2 \circ f_1) \otimes (g_2 \circ g_1) = (f_2 \otimes g_2) \circ (f_1 \otimes g_1) \tag{31}$$

becomes self-evident, as when building the diagrams corresponding to the left hand side and the right hand side of the formula above, one obtains the same diagram, namely,



. (32)

The second advantage is one of readability: when there are many systems interacting, it can start to become difficult to read the formulas, while the diagrams remain lucid.

Exercise II.1. Draw the diagram corresponding to the right hand side of (18).

B. Probabilistic interpretation

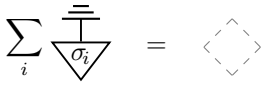
By adding extra conditions on a GPT, one can interpret every scalar as a probability. In some GPTs, there is a distinguished effect, called the *discard*, represented as


(33)

which represents ignoring the system from that point onwards. This allows to define probabilities in the following way. Let us introduce the empty diagram


(34)

to represent the number 1. Then a set of states $\{\sigma_i\}$ such that


(35)

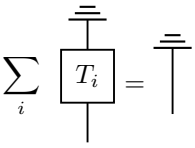
represents a preparation with a probabilistic outcome, where the label i serves to identify the different possible outcomes. Then the probability of outcome i happening is given by

$P\left(i|\{\sigma_i\}\right) =$  . (36)

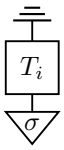
A *normalised* state, is one such that


(37)

Similarly, an *operation*, also known as an *instrument* is a set of transformations $\{T_i\}$ such that


(38)

and the probability of the particular transformation T_i happening, given that the system was in the normalised state σ is

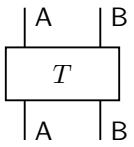
$P\left(i|\{T_i\},\sigma\right) =$  (39)

With this probabilistic structure, the requirement of convexity is understood, as in the first part of the lecture, as the requirement that, given the experimental ability of preparing the state σ and the ability of preparing the state σ', one is also able to flip a biased coin and prepare the state σ with probability p or σ' with probability $1-p$.

C. Causality and the conservation of probabilities

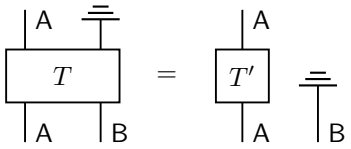
As we have just seen, the discard is closely related to the idea that probabilities always sum to 1. However, there is a close relation in GPTs between the conservation of probabilities and notions of *causality*. Indeed, diagrams that only make use of deterministic processes to propagate information, two processes can affect each other only if there is a system connecting them. Moreover, the result of an operation cannot be affected by the nature of a *later* operation. Let us see why.

A transformation T' on two systems A and B



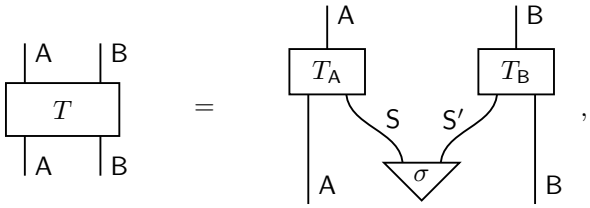
(40)

is said to be *non-signalling from* A *to* B if



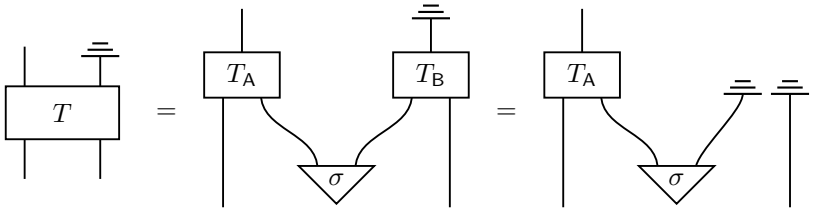
(41)

for some T', meaning that if one ignores B after applying T, then one can compute the resulting transformation on A without knowing the initial state of B. Put in other words, the initial state of B does not affect the statistics of A via T. It is immediate to prove that any transformation of the form:



(42)

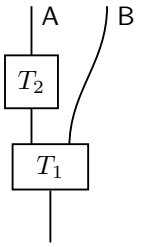
where $T_\mathsf{A}, T_\mathsf{B}$ and σ are normalised, is non-signalling from A to B and from B to A. Indeed



(43)

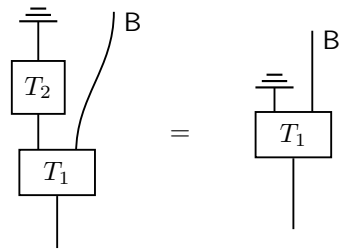
Note that this is quite a general statement. Indeed, σ could be a quantum entangled state, or a state of a theory that allows even stronger correlations. The only substantial requirement is that S and S′ are GPT systems.

Similarly, consider a diagram of the form



(44)

where both T_1 and T_2 are normalised. Then to compute the effect on system B, there is no need to know anything about T_2, since



(45)

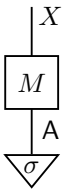
Thus, in GPTs, later operations do not affect the outcomes of earlier operations. This property is also called *no-signalling from the future.*

D. Classical and non-classical systems

A suitable GPT model of a finite-dimensional classical system can be constructed as follows. Let X be the finite set of configurations of the system. Then the states of the GPT are the probability distributions over X, the transformations are stochastic maps on these distributions, and the only effect (the discard) is marginalisation. The state space can be embedded in a $|X|$-dimensional real vector space V_X, the stochastic transformations are then represented by stochastic matrices and the discard amounts to summing all entries in the vector. Classical systems may be combined by making use of the tensor product of the underlying vector spaces.

The simplest nontrivial example of a classical system is the classical bit. It is associated with the space $X = \{0, 1\}$ and its state space is isomorphic to the line segment $[0, 1]$. The simplest nontrivial quantum system is the qubit, whose state space is isomorphic to a ball in 3D space.

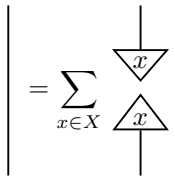
Classical systems can be used to model measurements on non-classical systems. Let M be a convex map from the states of A to states of a classical GPT system X, and let σ be a state for system A, then



(46)

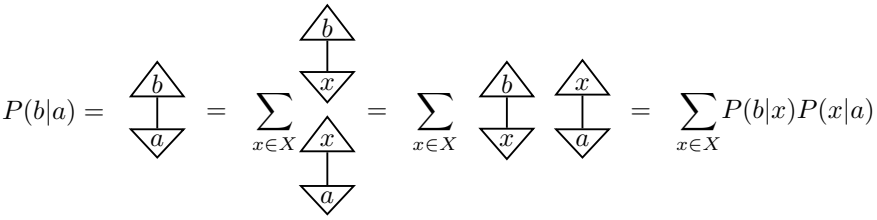
is a probability distribution over X, which can represent the probability of various outcomes of the measurement, as read on a classical pointer variable.

We can diagrammatically express what makes a system classical or not. For example, all classical systems have a crucial property, namely, that the identity operation can be decomposed as a sum of (or integral over) projectors:



(47)

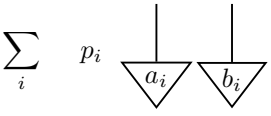
This property is also known as *atomicity of the identity*. Thus, when computing probabilities about classical systems, one can use the classical probability axiom



(48)

This can be taken as meaning that a classical system is always in one of its states, and all probabilistic considerations are a result of ignorance. Or it can be taken to say that classical systems can be measured without perturbing the state. Or yet again, there are no interference effects in classical systems. Either way, this property does not hold in a general GPT, and in particular, it does not hold in quantum theory.

Another property shared by all classical systems is that of *state separability*. A bipartite state is *separable* if it can be written as a convex combination of factored states:



(49)

All bipartite classical systems only have separable states. This follows easily from the defining property (47). Thus, if a GPT system has non-separable states, then the system cannot be a classical system. This allows us to define entanglement beyond quantum theory:

Definition II.1 (GPT entanglement)**.** *Two GPT systems are* entangled *if they are not in a separable state.*

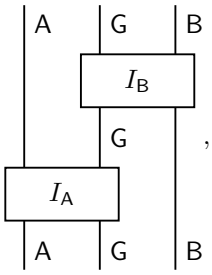
Exercise II.2. Prove that all states of a bipartite classical GPT system are in separable. Use the defining property of classicality (47).

III. NO-GO THEOREM ABOUT THE GRAVITATIONAL FIELD

We have now set up the machinery needed to understand the proof of the no-go theorem by Galley, Giacomini, and Selby [**?**].

We have the definition II.1 of entanglement for GTPs, we have the defining property of classicality (47), what we are left with is the notion of *mediation*.

Definition III.1 (Mediation)**.** *A system* G *is said to* mediate *the interaction between two systems* A *and* B *if the evolution can be written as*

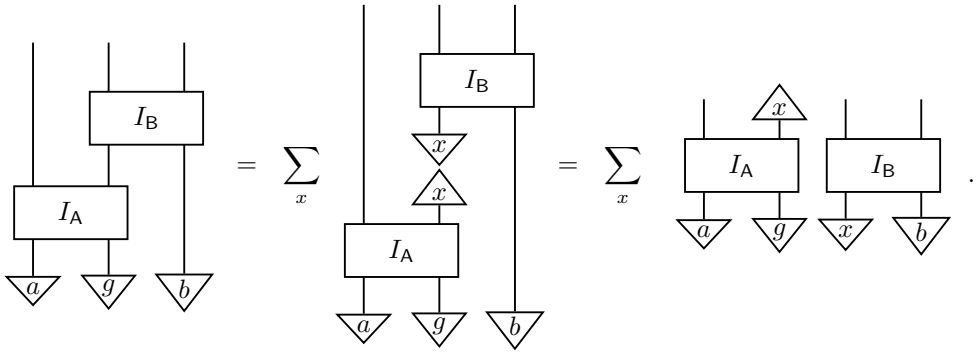


(50)

where I_A *and* I_B *are normalised, or as a sequence of such maps.*

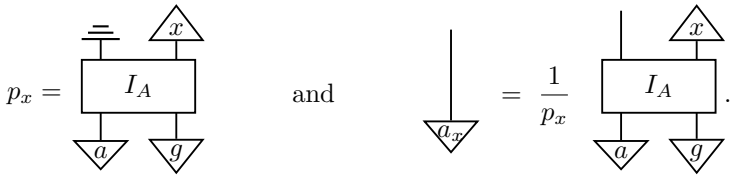
We are finally ready to prove the theorem.

Proof of theorem 1. Start by assuming that the three systems start in a separable state, and that the evolution is of the mediated form, with G classical. Using the defining property of classicality,



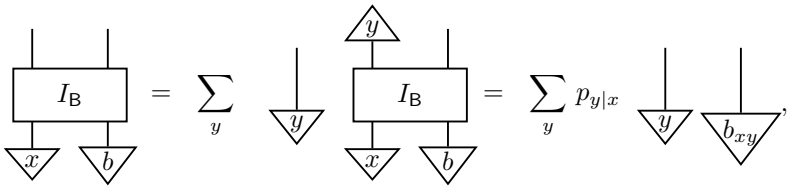
(51)

Then define



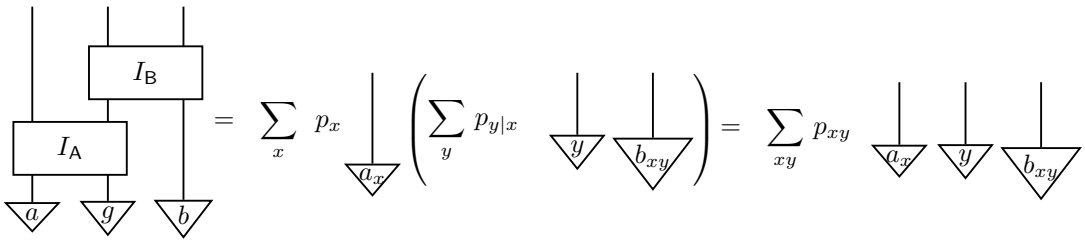
(52)

Since I_A, a and g are all normalised, we have $\sum_x p_x = 1$, and thus each a_x is also normalised. Thus (51) already shows that A is not entangled with B and G. To show that B and G are not entangled, we use again the defining property (47) of classicality



(53)

where we have defined the probability distribution $p_{y|x}$ and the normalised states b_{xy} in a manner analogous p_x and a_x. Putting all together,



(54)

The equation above shows clearly that the result of such interactions mediated by the classical system G leads to correlations, but not entanglement between A and B, as the final state is separable. This will be true of a sequence of such interactions. This completes the proof. □

Exercise III.1. Read the main text of [1] and the proof of the main theorem in the appendix.

IV. BONUS: ANOTHER NO-GO THEOREM ABOUT THE GRAVITATIONAL FIELD

Recently, Galley, Giacomini, and Selby published another theorem about the non-classicality of the gravitational field [17]:

Theorem 2 (No-Go Theorem 2)**.** Let A be a GPT system interacting with the gravitational field G via an interaction R. Then the following statements are incompatible:

- (Reversibility) The interaction R is reversible,
- (Information flow) There is information flow from A to G,
- (Non-classicality of A) A is fully nonclassical,
- (Classicality of G) G is a classical system.

This theorem generalises several arguments that have been presented in the literature that there is no consistent way to reversibly couple an actual quantum system to a classical system. If we have time, in the lecture, I will go over the concepts involved in this one.

Exercise IV.1. Read the main text of [17] and the proof of the main theorem in the appendix. Compare the diagrammatic with the algebraic proof.

---

[1] T. D. Galley, F. Giacomini, and J. H. Selby, A no-go theorem on the nature of the gravitational field beyond quantum theory, Quantum **6**, 779 (2022).

[2] S. Bose, A. Mazumdar, G. W. Morley, H. Ulbricht, M. Toroš, M. Paternostro, A. Geraci, P. Barker, M. S. Kim, and G. Milburn, A Spin Entanglement Witness for Quantum Gravity, Physical Review Letters **119**, 240401 (2017), arXiv:1707.06050.

[3] C. Marletto and V. Vedral, Gravitationally-induced entanglement between two massive particles is sufficient evidence of quantum effects in gravity, Physical Review Letters **119**, 240402 (2017), arXiv:1707.06036.

[4] A. Cabello, Interpretations of quantum theory: A map of madness, in *What Is Quantum Information?*, edited by O. Lombardi, S. Fortin, F. Holik, and C. López (Cambridge University Press, 2017) arXiv:1509.04711.

[5] C. A. Fuchs and A. Peres, Quantum Theory Needs No 'Interpretation', Physics Today **53**, 70 (2007).

[6] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, 10th ed. (Cambridge University Press, Cambridge ; New York, 2010).

[7] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Quantum entanglement, Reviews of Modern Physics **81**, 865 (2009), arXiv:quant-ph/0702225.

[8] A. Peres, Separability Criterion for Density Matrices, Physical Review Letters **77**, 1413 (1996), arXiv:quant-ph/9604005.
[9] M. Horodecki, P. Horodecki, and R. Horodecki, Separability of Mixed States: Necessary and Sufficient Conditions, Physics Letters A **223**, 1 (1996), arXiv:quant-ph/9605038.
[10] B. M. Terhal, Bell Inequalities and the Separability Criterion, Physics Letters A **271**, 319 (2000), arXiv:quant-ph/9911057.
[11] C. Marletto and V. Vedral, Witnessing non-classicality beyond quantum theory, Physical Review D **102**, 086012 (2020), arXiv:2003.07974.
[12] J. Barrett, Information processing in generalized probabilistic theories, Physical Review A **75**, 032304 (2007), arXiv:quant-ph/0508211.
[13] G. M. D'Ariano, G. Chiribella, and P. Perinotti, *Quantum Theory from First Principles: An Informational Approach* (Cambridge University Press, Cambridge, 2017).
[14] B. Coecke and A. Kissinger, *Picturing Quantum Processes* (Cambridge University Press, West Nyack, 2017).
[15] C. Marletto, Constructor Theory of Probability, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **472**, 20150883 (2016), arXiv:1507.03287.
[16] M. P. Müller, Probabilistic Theories and Reconstructions of Quantum Theory (Les Houches 2019 lecture notes), SciPost Physics Lecture Notes , 28 (2021), arXiv:2011.01286.
[17] T. D. Galley, F. Giacomini, and J. H. Selby, Any consistent coupling between classical gravity and quantum matter is fundamentally irreversible, Quantum **7**, 1142 (2023).