Any consistent coupling between classical gravity and quantum matter is fundamentally irreversible

Thomas D. Galley[1], Flaminia Giacomini[2], and John H. Selby[3]

[1]Institute for Quantum Optics and Quantum Information, Austrian Academy of Sciences, Boltzmanngasse 3, 1090 Vienna, Austria

[2]Institute for Theoretical Physics, ETH Zürich, 8093 Zürich, Switzerland

[3]ICTQT, University of Gdańsk, Wita Stwosza 63, 80-308 Gdańsk, Poland

When gravity is sourced by a quantum system, there is tension between its role as the mediator of a fundamental interaction, which is expected to acquire nonclassical features, and its role in determining the properties of spacetime, which is inherently classical. Fundamentally, this tension should result in breaking one of the fundamental principles of quantum theory or general relativity, but it is usually hard to assess which one without resorting to a specific model. Here, we answer this question in a theory-independent way using General Probabilistic Theories (GPTs). We consider the interactions of the gravitational field with a single matter system, and derive a no-go theorem showing that when gravity is classical at least one of the following assumptions needs to be violated: (i) Matter degrees of freedom are described by fully non-classical degrees of freedom; (ii) Interactions between matter degrees of freedom and the gravitational field are reversible; (iii) Matter degrees of freedom back-react on the gravitational field. We argue that this implies that theories of classical gravity and quantum matter must be fundamentally irreversible, as is the case in the recent model of Oppenheim et al. Conversely if we require that the interaction between quantum matter and the gravitational field is reversible, then the gravitational field must be non-classical.

1 Introduction

The gravitational field plays two roles. On the one hand, gravity is a fundamental interaction coupled to quantum matter, and as such it is natural to look for its quantum description. On the other hand, the gravitational field characterises the structure of spacetime, which is a classical concept coming from general relativity. Reconciling these two roles when the source of gravity is quantum is one of the most profound conceptual challenges in finding a unified theory of gravity and matter.

Recently, the possibility of testing the gravitational field of a quantum source has attracted a lot of attention [1–30]. No experiment using the technology currently available can test this scenario now [31], but these tests are expected to be feasible in the next couple of decades, and would bring a major advancement to our understanding of gravity in regimes so far fully unexplored. On the theoretical side, the precise implications of these experiments are still debated [24]. From a fundamental point of view, however, this regime gives us the opportunity to test the fundamental principles of both quantum theory and general relativity together in a single scenario. Hence, it is crucial now to assess what the logical implications of combining the two theories in this regime are.

Here, we set to this task, and we ask whether we need to give up any of the fundamental principles of our theories. To achieve this goal, we use a theory-independent approach, analogous to that used in Bell's theorem [32], via the framework of General Probabilistic Theories (GPTs)

Thomas D. Galley: thomas.galley@oeaw.ac.at
Flaminia Giacomini: fgiacomini@phys.ethz.ch
John H. Selby: john.h.selby@gmail.com

[33, 34] in contrast to earlier work which tends to consider specific, theory-dependent, models of interactions between matter and gravity, such as the Schrödinger-Newton equation. In this theory independent approach, we prove a no-go theorem, in the spirit of our recent work of Ref. [18], that states that, if we want to keep a classical description of gravity and we allow quantum matter to source the gravitational field, then it is impossible to preserve the principle of reversibility of fundamental interactions. In general, our theorem shows under which conditions non-classical (including quantum) and classical degrees of freedom can couple in a consistent way. This is more general than previous work which was restricted to studying classical-quantum coupling [35–39].

We formally state the no-go theorem after introducing the relevant preliminaries of the theory independent framework which we work in. Informally however, the no-go theorem states that given a general classical system (in our case, the gravitational field), then a physical system coupled to it cannot at once (i) satisfy the superposition principle, (ii) interact reversibly with the classical system, and (iii) back react on the classical system. Formally, the theorem holds for finite-dimensional systems. However, operationally every physical system can be reduced to a finite dimensional system, because in the laboratory we can only perform a finite set of measurements.

From our result, it is clear that if one wants to preserve these three principles, then the gravitational field ought to be nonclassical. Otherwise, if one wants to hold to classical gravity, then the only consistent ways to couple a quantum system with a classical system such that there is back-reaction from the quantum to the classical system are to either a) have an irreversible interaction as in Refs. [40–42] or b) have a fundamentally superselected quantum system with only the classical "which sector" degrees of freedom interacting with the classical system.

In the GPT literature similar theorems have been proved to study measurement theory [43]. In this case the classical system is a measuring device and the system being measured is non-classical. Since a required feature of a measurement interaction is that the state of the measuring device after the interaction is a function of the state of the system there must be back-reaction. Thus measurement of non-classical systems must be irreversible (i.e., involve disturbance in the language of [43]).

2 General probabilistic theories: a theory independent framework

A generalised probabilistic theory is a mathematical object which precisely captures the operationally relevant content of a physical theory [44], at least for scenarios involving single agents and a fixed causal structure. That is, it captures only the information in the physical theory which is necessary and sufficient for making predictions about the outcomes of any such experiments that can be performed within the theory. Operationally, the state of a system is characterised by finitely many measurements, and hence the state space of the system is finite dimensional. This entails that quantum systems have finite dimensional Hilbert spaces and classical systems have finite configuration spaces. Alternatively one could accept the infinite dimensional nature of systems such as the gravitational field and instead understand the finite dimensional GPT description as a coarse grained version of the underlying infinite dimensional system.

A given physical theory may of course do more than describing the operational aspects, it may also describe ontological or interpretational aspects. For example, a physical theory could be Bohmian mechanics [45] or Many-Worlds quantum theory [46], but both would be associated to the same generalised probabilistic theory, namely, the standard formalism of quantum (information) theory. It is therefore natural to assume that, whatever the underlying theory of nature actually is, there is a regime in which it is well described by a GPT[1]. This regime at least includes all of the standard kinds of experiments which we perform, but, the particular formalism that we use here would exclude for example,

[1]It is, however, often highly nontrivial to determine which GPT is associated to a given physical theory and there are often surprising results. For example, the GPT associated to quantum theory with a non-linear modification of the Schrödinger equation is actually a classical GPT [47], as the non-linearity allows for perfect distinguishability of all pure states. For some relevant models, such as the Reginatto-Hall model [48] the GPT description is not known, and as such our results do not immediately apply.

experiments involving indefinite causal structure [49–51] (in which the pieces of apparatus locally have a well defined causal structure but in which no global causal structure exists between them) or extended Wigner's friend scenarios [52–57] (in which the experimenters themselves are treated as part of the experimental scenario). The GPT framework that we present here can, however, also be extended to include these more general scenarios, for example [49, 58–62]. Any results that can be proven in this bare-bones operational framework therefore automatically apply to most physical theories even if they are not themselves expressed in the language of GPTs.

More concretely, a GPT describes a set of preparations, transformations and measurements and encodes the probabilities of obtaining measurement outcomes given a circuit of preparations, transformations and measurements. These probabilities may be given by classical or quantum theory, or some different theory. In this section we do not provide a full introduction to GPTs but rather refer the reader to the works [63–65]. Here, we introduce the notation and important definitions required for the main theorem.

In the following we denote by Ω the convex set of states of a GPT system S, $\mathcal{E}$ the set of effects and $\mathcal{T}$ the set of transformations. If V is the finite dimensional real vector space associated to S then $\Omega \subset V$[2], $\mathcal{E} \subset V^*$ and any $t \in \mathcal{T}$ induces a linear transformation on V. In a given GPT two systems S_1 and S_2 with associated vector spaces V_1 and V_2 compose to a third system S_3 with associated vector space V_3 via a bilinear map $\otimes : V_1, V_2 \to V_3$. When the theory is locally tomographic [33] (such as is the case for quantum and classical theory) $\otimes$ is the standard tensor product.

Definition 1 (Reversible interaction)**.** A reversible interaction between two GPT systems S_1 and S_2 with respective state spaces Ω_1 and Ω_2 is a linear map $\mathcal{I} : \Omega_1 \otimes \Omega_2 \to \Omega_1 \otimes \Omega_2$ where there exists an inverse map $\mathcal{I}^{-1} : \Omega_1 \otimes \Omega_2 \to \Omega_1 \otimes \Omega_2$ such that $\mathcal{I}^{-1} \circ \mathcal{I} = \mathbb{I}_{\Omega_1 \otimes \Omega_2}$.

Definition 2 (Information flow)**.** Given an interaction $\mathcal{I}$ between two GPT systems S_1 and S_2 with input states s_1 and s_2 there is information flow from S_1 to S_2 if a change in state of system S_1 before the interaction can lead to a change of the state of system S_2 after the interaction.

If we denote the choices of state preparation by the classical variables s_1 and s_2 and denote the outcomes of possible measurements after the interaction by the variables e_1 and e_2 this is equivalent to saying there is signalling from 1 to 2, that is, $p(e_2|s_1, s_2, \mathcal{I})$ depends non-trivially on s_1.

Definition 3 (Finite dimensional classical system)**.** A classical system of dimension n has a state space given by an n-vertex simplex Δ_n with associated vector space $V \cong \mathbb{R}^{n+1}$. Effects belong to the hypercube of $[0, 1]$-valued linear functionals in $V^* \cong \mathbb{R}^{n+1}$ and transformations are given by stochastic linear maps. Classical systems compose with all other systems (both classical and non-classical) via the tensor product.

Intuitively, this definition states that any two classical states can be perfectly distinguished with an operation belonging to the set of classical effects, and that any combination of two classical states leads at most to a classical mixture of probabilities (e.g., no quantum interference is allowed). Finally, the standard tensor product composition rule ensures that no entangled states are generated when two classical systems are combined.

Definition 4 (Reducible GPT system)**.** A GPT system S, with associated vector space V, is reducible if and only if there is a decomposition $V \cong \bigoplus_i V_i$ such that all states are convex combinations of states having support in a single block. That is, such that the state space $\Omega \cong \bigoplus_i \Omega_i$ where Ω_i is a convex set in V_i.

A reducible GPT system is one that has a classical degree of freedom, and, hence, is at least partly classical. For example, the finite dimensional classical systems described above can be written as $\Delta_n = \underbrace{\Delta_1 \oplus \cdots \oplus \Delta_1}_{n}$ and, more generally, if we have a reducible system of the form $\underbrace{\Omega \oplus \cdots \oplus \Omega}_{n}$ then this is nothing but $\Omega \otimes \Delta_n$, that is, it factorises as a classical system, Δ_n and another system Ω. Finally, any reducible system $\bigoplus_{i=1}^n \Omega_i$ can be thought of as a probabilistic mixture of the systems Ω_i, where there is some label i telling us which system we have. This label cannot be prepared in 'superpositions' and so is

[2] Where the symbol $\subset$ means "is a convex subset of".

best thought of as a classical degree of freedom. In the case of quantum theory, reducible systems are those that have a non-trivial superselection rule, and the classical label tells us which superselection sector the state has been prepared in. This motivates the following definition:

Definition 5 (Fully nonclassical systems)**.** A fully nonclassical GPT system S is one that is irreducible.

Irreducible systems can be characterised by the fact that their identity transformations cannot be viewed as a coarse graining of other transformations, or, equivalently [66], that there do not exist any non-disturbing measurements for the system. See also [43, Section 5.1] for more discussion of reducible GPT systems.

3 Results

Our main technical result is the following no-go theorem, which follows directly from Theorem 2 proven in Appendix D. Moreover an algebraic version of the proof for the special case of quantum systems is given in Appendix E.1 for the reader who is less familiar with diagrammatic approaches.

Theorem 1. Given two GPT systems S and G which interact via some interaction $\mathcal{I}$ then at least one of the following conditions must be violated:

- (i) The system S is fully non-classical (Def. 5)
- (ii) The interaction $\mathcal{I}$ is reversible (Def. 1)
- (iii) There is information flow from system S to system G (Def. 2)
- (iv) G is classical (Def. 3)

Furthermore Theorem 2 shows that if S is reducible and G is classical, it is only the classical (i.e., the which-sector) information which can flow from S to a G via a reversible interaction $\mathcal{I}$.

Proof sketch: Suppose that conditions $(i)-(iv)$ of the theorem hold and consider the following procedure. First evolve G and S by $\mathcal{I}$, which by condition (iii) constitutes a non-trivial measurement of S (see Example 2). Following this the state of G state can be copied onto another system G' (in the sense of Equation (41)) since G is classical by condition (iv). Subsequently the inverse $\mathcal{I}^{-1}$ evolution (which exists by condition (ii)) is applied which returns S and G to their initial states. The information gained about S during $\mathcal{I}$ is not erased since it has been copied onto G'. Hence this total procedure constitutes a disturbance free measurement of the irreducible system S (condition (i)) with non-trivial information gain. It follows from the information gain versus state disturbance trade-offs that this is impossible and therefore at least one of the assumptions $(i)-(iv)$ must be violated.

We can intuitively understand this theorem in the case of quantum theory by appealing to an example from Ref. [37]. Consider a classical harmonic oscillator (system G) with free Hamiltonian $H_0 = \frac{1}{2}p^2$ interacting with a quantum spin system (system S) via the interaction Hamiltonian $H_I = \kappa\hat{\sigma}_3 p$. The x coordinate cannot satisfy $\dot{x} = \partial_p(H_0+H_I) = p+\kappa\hat{\sigma}_3$, where the dot denotes the time derivative and ∂_p stands for the partial derivative by p, since $\hat{\sigma}_3$ is an operator. Hence we might use the expectation value $\langle\hat{\sigma}_3\rangle$ instead. This is then an interaction which is reversible and admits back-reaction, hence obeys conditions (ii) and (iii). However, this interaction

> *"implies that quantum expectations can be deduced with arbitrary precision from the measurement of the classical variables x and p."* [37].

This means that either the classical system must inherit some of the quantum uncertainty (and no longer be classical) in order to prevent violation of the uncertainty principle of the quantum system or that the quantum system no longer satisfies the uncertainty principle, and so has inherited classical properties. In other words either condition (iv) is violated since G has become nonclassical, or condition (i) is violated, since S has become classical.

Let us emphasise that the proof of the theorem only requires us to consider interactions between a classical system and a non-classical system – in particular, we do not need any assumption on the properties of a nonclassical theory of gravity (e.g., Hilbert space factorisation, interactions, etc.).

3.1 Examples

We provide some simple examples of different types of interactions between systems and which of the assumptions of the theorem they meet.

Example 1 (Entangling description of a quantum measurement)**.** Consider two qubits S and G evolving according to a CNOT gate. This constitutes a Z measurement of S by G when G is initialised in $|0\rangle$:

$$(\alpha |0\rangle_{\mathsf{S}} + \beta |1\rangle_{\mathsf{S}}) |0\rangle_{\mathsf{G}} \mapsto \alpha |00\rangle_{\mathsf{SG}} + \beta |11\rangle_{\mathsf{SG}} \quad (1)$$

Here, conditions (i), (ii) and (iii) are met while condition (iv) is not. To see that condition (iii) is met observe that the final state $|\alpha|^2 |0\rangle\langle 0|_{\mathsf{G}} + |\beta|^2 |1\rangle\langle 1|_{\mathsf{G}}$ of G depends on the initial state $\alpha |0\rangle_{\mathsf{S}} + \beta |1\rangle_{\mathsf{S}}$ of S.

Example 2 (Measurement of a quantum system by a classical system)**.** We now consider a qubit S on which a classical two level system G (represented by a superselected qubit in the Z basis) performs a Z observable measurement. The two level system evolves according a decohered CNOT gate:

$$|\psi\rangle\langle\psi| \otimes |0\rangle\langle 0| \mapsto |\alpha|^2 |00\rangle\langle 00|_{\mathsf{SG}} + |\beta|^2 |11\rangle\langle 11|_{\mathsf{SG}}, \quad (2)$$

where $|\psi\rangle = \alpha |0\rangle_{\mathsf{S}} + \beta |1\rangle_{\mathsf{S}}$. Here, conditions (i), (iii) and (iv) are met whilst condition (ii) is not.

Example 3 (Reversible measurement of superselection sector by classical systems)**.** Let us consider two spin $\frac{1}{2}$ systems with a total angular momentum super-selection rule; i.e. where there cannot be coherence between sectors with different total angular momentum. Using known decomposition rules for total angular momentum we have:

$$\frac{1}{2} \otimes \frac{1}{2} \simeq 0 \oplus 1, \quad (3)$$

where the Hilbert space decomposes into $\mathbb{C}^1 \oplus \mathbb{C}^3$ corresponding to the anti-symmetric and symmetric subspaces of $\mathbb{C}^2 \otimes \mathbb{C}^2$. A general density operator for this super-selected system is block diagonal $\rho = \rho_0 \oplus \rho_1$. We can define the following total angular momentum (i.e. "which sector") measurement by a classical two level system:

$$(\rho_0 \oplus \rho_1) \otimes |0\rangle\langle 0| \mapsto (\rho_0 \otimes |0\rangle\langle 0|) \oplus (\rho_1 \otimes |1\rangle\langle 1|) \quad (4)$$

The above example is a specific instance of superselection sectors induced by the action of a group G on a Hilbert space $\mathcal{H}$ [67]. Here, conditions (ii), (iii) and (iv) are met whilst condition (i) is not.

Example 4 (Semi-classical gravity (see Appendix B))**.** In semi-classical gravity the stress energy tensor $T_{\mu\nu}$ in the Einstein equation is replaced by the expectation value of the stress energy tensor $\langle T_{\mu\nu}\rangle_{|\psi\rangle}$ for the matter field $|\psi\rangle$:

$$R_{\mu\nu} + \frac{1}{2} g_{\mu\nu} R = \frac{8\pi G}{c^4} \langle\psi| \hat{T}_{\mu\nu} |\psi\rangle . \quad (5)$$

From this it follows (under some additional assumptions) that $|\psi\rangle$ evolves according to the (non-linear) Schrödinger-Newton equation [68]:

$$i\frac{d}{dt}\psi = \left[-\frac{\nabla^2}{2m} - Gm^2 \int d^3r' \frac{|\psi(t,r')|^2}{|r-r'|} + V \right] \psi . \quad (6)$$

The key feature of semi-classical gravity (and semi-classical approximations more generally as described in Appendix A) is that they induce dynamics described by a non-linear Schrödinger equation. Importantly 'quantum' systems with such modified dynamics no longer have the same operational predictions as quantum theory, and as such do not correspond to the same GPT. It was shown in Ref. [47] for the family of NLSE which are non-linear in $|\psi|^2$ that such systems are in fact fully classical, with states corresponding to probability measures over the projective Hilbert space of quantum states. This is analogous to the 'classicalisation' of the quantum spin system coupled reversibly to the classical harmonic oscillator discussed above. This then shows that semi-classical gravity, which claims to couple quantum matter with a classical gravitational field, violates condition (i) and therefore describes classical matter interacting with a classical gravitational field.

In semi-classical gravity the Schödinger-Newton equation appears as a *fundamental* description of quantum matter interacting with classical gravity [68]. We note that non-linear Schrödinger equations, including the Schrödinger-Newton equation also appear in effective descriptions of quantum systems via mean-field approximations [68, 69]. We discuss these in Appendix A and note that whilst the theorem also applies to them, they are of less conceptual interest since they are understood to be useful approximations for describing interacting quantum systems rather than a fundamental description of quantum-classical interaction.

4 Discussion

Let us discuss what a rejection of each of the conditions in Theorem 1 entails:

Condition (i). Rejecting this condition would mean that any matter degrees of freedom which have a back-action on the gravitational field (for example the position of a massive particle) are modelled by reducible systems, for instance superselected quantum systems or classical systems. Since massive quantum particles are irreducible systems, this requires us to reject quantum theory as describing any matter degrees of freedom which back-react on the gravitational field. Essentially, rejecting condition (i) means rejecting quantum theory and the superposition principle. We note that there are two main ways in which a quantum system can 'classicalise'. The first is via a superselection rule, so that superpositions of states in a given basis are not allowed. The second is that the projective Hilbert space of pure quantum states is in fact a classical configuration space and the mixed states are measures on this space. For a qubit which has pure states which are isomorphic to a sphere, the corresponding classical system consists of all measures on the sphere with pure states being Dirac measures [3]. In this case a pure state $\delta_{\frac{1}{\sqrt{2}}(|0\rangle+|1\rangle)}$ is perfectly distinguishable from the states $\delta_{|0\rangle}$ and $\delta_{|1\rangle}$. This is the case, e.g., of the Schrödinger-Newton equation, discussed in Appendix B.

Condition (ii). Rejecting this condition can lead to the introduction of some sort of stochasticity, as in Refs. [40, 41, 71]. For instance, in [40, 41] a model of quantum matter coupled to classical gravity is proposed. The interaction leads to linear dynamics in the density matrix, unlike proposals such as the Schrödinger-Newton equation for instance, meaning that the matter degrees of freedom are fully described by quantum theory and therefore non-classical. The evolution is stochastic and is equivalent to an open quantum system description for the matter degrees of freedom. There is back-reaction of the matter degrees of freedom on the gravitational field. As such this proposal violates condition (ii), whilst meeting conditions (i) and (iii).

[3]This classical system appears in the Beltrametti-Bugajski model of quantum systems [70].

Condition (iii). The case where there is no back-reaction from matter onto the gravitational field may emerge as a good approximation in certain situations, such as a low mass particle interacting gravitationally with a large static body. However, while condition (iii) can be violated in certain approximations it would be in contradiction with a fundamental feature of gravity if the gravitational field was not influenced by matter.

Condition (iv). In this case the gravitational field is non-classical. In the context of the current discussion on the low-energy nature of the gravitational field [1–30], the violation of this condition is the only one allowing one to preserve both reversibility and back-action when gravity interacts with quantum matter. The violation of this condition thus consists in a justification of the search for an (unknown) quantum gravity theory at higher energies, which might have very different features to those predicted in the low-energy regime. Whilst non-classicality of the gravitational field is consistent with the interaction being either reversible or irreversible we note that all the proposals the authors are aware of are reversible.

4.1 On reversibility as a feature of quantum matter interacting with the gravitational field

Let us now apply the theorem to the question of whether the gravitational field needs to be quantized. We assume that matter is irreducibly quantum (condition (i)) and that it back-reacts on the gravitational field (condition iii).

If one is committed to reversibility as a fundamental feature of the interaction between matter and the gravitational field, then one is forced to conclude that the gravitational field is non-classical. If one is not committed to reversibility then this leaves open the possibility of the gravitational field being classical.

Theorem 1 is not normative and gives no reason to prefer rejecting one assumption over another. Rather one must appeal to external arguments to justify different choices. There is ample discussion about how fundamental reversibility is in physics. On the one hand, quantum field theory (our best fundamental microscopic theory of matter) and general relativity (our best theory of the interaction between classical matter and space-time) are reversible. Hence, one may expect that a theory combining the two should preserve reversibility. On the other hand, we have potential instances of irreversibility in fundamental physical theories, such as measurements in quantum theory and thermodynamical processes. Even though one may argue that at a more fundamental level thermodynamics can be reduced to statistical mechanics, this reduction is controversial and the status of reversibility within statistical mechanics is still debated.

Finally, we observe that reversibility is not straightforward to falsify, since the observation of an irreversible process does not preclude the possibility that it is a part of a larger reversible process.

4.2 Should the gravitational field be quantized?

When studying reversible interactions between quantum matter and classical gravity Mielnik concluded that "Either gravity is quantum or the electron is classical" and therefore "though it may be very difficult to quantize the gravitation, it is even more difficult not to do it" [72]. Similarly [5] states that "everything must be quantized" using the framework of constructor theory. This framework assumes reversibility as well as the assumption that 'every measurer of a Z observable' (which is the CNOT gate of Ex.1 for quantum theory) can be used to perfectly discriminate states in the X basis (which is met in quantum theory since the input states $|\pm\rangle\,|0\rangle$ are mapped to orthogonal states). Moreover superselected systems are not considered since the non-classical system is assumed to have just two observables, which are incompatible While both these results [5, 72] form part of a larger debate on the necessity of quantizing the gravitational field, the present work makes precise exactly which assumptions are needed to construct such arguments, for instance by highlighting the key roles reversibility and superselection play.

The results of this paper show that the above statements about the necessity of quantizing the gravitational field are only true under the assumption of reversibility, and that moreover if one accepts super-selected quantum matter one can preserve both reversibility and classicality of the gravitational field.

A natural extension of the results of this paper is to consider infinite dimensional systems. For an extension to infinite dimensional classical systems interacting with irreducible finite dimensional quantum systems the algebraic version of the proof in Appendix E.1 can informally be adapted by replacing $\sum_x \to \int_x dx$. The diagrammatic proof of Appendix D is expected to extend to the general case, however a fully rigorous extension to infinite dimensional classical and non-classical systems is left to future work.

4.3 Classical gravity and quantum matter

If gravity is classical, at least one of the conditions (i)-(iii) has to fail. In summary, this implies the following.

Fail (i)	Reject irreducibly quantum matter
Fail (ii)	Reject reversibility
Fail (iii)	Reject matter dependence of gravity

Hence, a theory in which classical general relativity holds and matter is quantum necessarily needs to break the principle of reversibility. This means that the main feature of the model of [40, 41] which one could object to, namely stochasticity of the dynamics, is in fact inevitable for any model which seeks to combine quantum matter with classical general relativity.

Acknowledgments

The authors would like to thank Barbara Šoda for helpful feedback on the manuscript and the referees for their comments which helped improve the manuscript. T.D.G acknowledges support from the Austrian Science Fund (FWF), project P 33730-N. F.G. acknowledges support from the Swiss National Science Foundation via the Ambizione Grant PZ00P2-208885, from the ETH Zurich Quantum Center, and from the John Templeton Foundation, as part of the 'The Quantum Information Structure of Spacetime, Second Phase (QISS 2)' Project. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation. T.D.G. and F.G. acknowledge support from Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is supported in part by the Government of Canada through the Department of Innovation, Science and Economic Development and by the Province of Ontario through the Ministry of Colleges and Universities. J.H.S. was supported by the Foundation for Polish Science through IRAP project co-financed by EU within Smart Growth Operational Programme (contract no. 2018/MAB/5).

Author contributions: All authors contributed equally to this work.

References

[1] M Bahrami, A Bassi, S McMillen, M Paternostro, and H Ulbricht. "Is gravity quantum?" (2015). arXiv:1507.05733.

[2] Charis Anastopoulos and Bei-Lok Hu. "Probing a gravitational cat state". Class. Quant. Grav. **32**, 165022 (2015).

[3] Sougato Bose, Anupam Mazumdar, Gavin W Morley, Hendrik Ulbricht, Marko Toroš, Mauro Paternostro, Andrew A Geraci, Peter F Barker, MS Kim, and Gerard Milburn. "Spin entanglement witness for quantum gravity". Phys. Rev. Lett. **119**, 240401 (2017).

[4] Chiara Marletto and Vlatko Vedral. "Gravitationally induced entanglement between two massive particles is sufficient evidence of quantum effects in gravity". Phys. Rev. Lett. **119**, 240402 (2017).

[5] Chiara Marletto and Vlatko Vedral. "Why we need to quantise everything, including gravity". npj Quantum Information **3**, 1–5 (2017).

[6] Matteo Carlesso, Mauro Paternostro, Hendrik Ulbricht, and Angelo Bassi. "When Cavendish meets Feynman: A quantum torsion balance for testing the quantumness of gravity" (2017). arXiv:1710.08695.

[7] Michael JW Hall and Marcel Reginatto. "On two recent proposals for witnessing nonclassical gravity". J. Phys. A **51**, 085303 (2018).

[8] Chiara Marletto and Vlatko Vedral. "When can gravity path-entangle two spatially superposed masses?". Phys. Rev. D **98**, 046001 (2018).

[9] Alessio Belenchia, Robert M Wald, Flaminia Giacomini, Esteban Castro-Ruiz, Časlav Brukner, and Markus Aspelmeyer. "Quantum superposition of massive objects and the quantization of gravity". Phys. Rev. D **98**, 126009 (2018).

[10] Alessio Belenchia, Robert M Wald, Flaminia Giacomini, Esteban Castro-Ruiz, Časlav Brukner, and Markus Aspelmeyer. "Information content of the gravitational field of a quantum superposition". Int. J. Mod. Phys. D **28**, 1943001 (2019).

[11] Marios Christodoulou and Carlo Rovelli. "On the possibility of laboratory evidence for quantum superposition of geometries". Phys. Lett. B **792**, 64–68 (2019).

[12] Charis Anastopoulos and Bei-Lok Hu. "Quantum superposition of two gravitational cat states". Class. Quant. Grav. **37**, 235012 (2020).

[13] Richard Howl, Vlatko Vedral, Devang Naik, Marios Christodoulou, Carlo Rovelli, and Aditya Iyer. "Non-gaussianity as a signature of a quantum theory of gravity". PRX Quantum **2**, 010325 (2021).

[14] Ryan J Marshman, Anupam Mazumdar, and Sougato Bose. "Locality and entanglement in table-top testing of the quantum nature of linearized gravity". Phys. Rev. A **101**, 052110 (2020).

[15] Hadrien Chevalier, A. J. Paige, and M. S. Kim. "Witnessing the nonclassical nature of gravity in the presence of unknown interactions". Phys. Rev. A **102**, 022428 (2020). arXiv:2005.13922.

[16] Tanjung Krisnanda, Guo Yao Tham, Mauro Paternostro, and Tomasz Paterek. "Observable quantum entanglement due to gravity". npj Quantum Information **6**, 1–6 (2020).

[17] Chiara Marletto and Vlatko Vedral. "Witnessing nonclassicality beyond quantum theory". Phys. Rev. D **102**, 086012 (2020).

[18] Thomas D. Galley, Flaminia Giacomini, and John H. Selby. "A no-go theorem on the nature of the gravitational field beyond quantum theory". Quantum **6**, 779 (2022).

[19] Soham Pal, Priya Batra, Tanjung Krisnanda, Tomasz Paterek, and T. S. Mahesh. "Experimental localisation of quantum entanglement through monitored classical mediator". Quantum **5**, 478 (2021).

[20] Daniel Carney, Holger Müller, and Jacob M. Taylor. "Using an Atom Interferometer to Infer Gravitational Entanglement Genera-

tion". PRX Quantum **2**, 030330 (2021). arXiv:2101.11629.
[21] Kirill Streltsov, Julen Simon Pedernales, and Martin Bodo Plenio. "On the significance of interferometric revivals for the fundamental description of gravity". Universe**8** (2022).
[22] Daine L. Danielson, Gautam Satishchandran, and Robert M. Wald. "Gravitationally mediated entanglement: Newtonian field versus gravitons". Phys. Rev. D **105**, 086001 (2022). arXiv:2112.10798.
[23] Adrian Kent and Damián Pitalúa-García. "Testing the nonclassicality of spacetime: What can we learn from Bell–Bose et al.-Marletto-Vedral experiments?". Phys. Rev. D **104**, 126030 (2021).
[24] Marios Christodoulou, Andrea Di Biagio, Markus Aspelmeyer, Časlav Brukner, Carlo Rovelli, and Richard Howl. "Locally mediated entanglement in linearized quantum gravity". Phys. Rev. Lett. **130**, 100202 (2023). arXiv:2202.03368.
[25] Nick Huggett, Niels Linnemann, and Mike Schneider. "Quantum Gravity in a Laboratory?" (2022). arXiv:2205.09013.
[26] Marios Christodoulou, Andrea Di Biagio, Richard Howl, and Carlo Rovelli. "Gravity entanglement, quantum reference systems, degrees of freedom" (2022). arXiv:2207.03138.
[27] Daine L. Danielson, Gautam Satishchandran, and Robert M. Wald. "Black Holes Decohere Quantum Superpositions" (2022). arXiv:2205.06279.
[28] Lin-Qing Chen, Flaminia Giacomini, and Carlo Rovelli. "Quantum states of fields for quantum split sources". Quantum **7**, 958 (2023). arXiv:2207.10592.
[29] Eduardo Martín-Martínez and T. Rick Perche. "What gravity mediated entanglement can really tell us about quantum gravity" (2022). arXiv:2208.09489.
[30] Chris Overstreet, Joseph Curti, Minjeong Kim, Peter Asenbaum, Mark A. Kasevich, and Flaminia Giacomini. "Inference of gravitational field superposition from quantum measurements" (2022). arXiv:2209.02214.
[31] Markus Aspelmeyer. "When Zeh Meets Feynman: How to Avoid the Appearance of a Classical World in Gravity Experiments". Fundam. Theor. Phys. **204**, 85–95 (2022). arXiv:2203.05587.
[32] John S Bell. "On the Einstein Podolsky Rosen paradox". Physics Physique Fizika **1**, 195 (1964).
[33] Lucien Hardy. "Quantum theory from five reasonable axioms" (2001). arXiv:quant-ph/0101012.
[34] Jonathan Barrett. "Information processing in generalized probabilistic theories". Physical Review A **75**, 032304 (2007).
[35] L. Diosi and J. J. Halliwell. "Coupling Classical and Quantum Variables using Continuous Quantum Measurement Theory". Physical Review Letters **81**, 2846–2849 (1998).
[36] J. Caro and L. L. Salcedo. "Impediments to mixing classical and quantum dynamics". Physical Review A **60**, 842–852 (1999).
[37] Lajos Diósi, Nicolas Gisin, and Walter T. Strunz. "Quantum approach to coupling classical and quantum dynamics". Physical Review A **61**, 022108 (2000).
[38] Daniel R. Terno. "Inconsistency of quantum–classical dynamics, and what it implies". Foundations of Physics **36**, 102–111 (2006).
[39] Hans-Thomas Elze. "Linear dynamics of quantum-classical hybrids". Physical Review A **85**, 052109 (2012).
[40] Jonathan Oppenheim. "A post-quantum theory of classical gravity?" (2018). arXiv:1811.03116.
[41] Jonathan Oppenheim, Carlo Sparaciari, Barbara Šoda, and Zachary Weller-Davies. "Gravitationally induced decoherence vs space-time diffusion: testing the quantum nature of gravity" (2022). arXiv:2203.01982.
[42] Isaac Layton, Jonathan Oppenheim, and Zachary Weller-Davies. "A healthier semi-classical dynamics" (2022). arXiv:2208.11722.
[43] Teiko Heinosaari, Leevi Leppäjärvi, and Martin Plávala. "No-free-information principle in general probabilistic theories". Quantum **3**, 157 (2019).
[44] Giulio Chiribella, Giacomo Mauro D'Ariano, and Paolo Perinotti. "Probabilistic theories with purification". Physical Review A **81**, 062348 (2010).
[45] David Bohm. "A suggested interpretation

of the quantum theory in terms of" hidden" variables. I". Physical review **85**, 166 (1952).

[46] Hugh Everett. "The theory of the universal wave function". In The many-worlds interpretation of quantum mechanics. Pages 1–140. Princeton University Press (2015).

[47] Bogdan Mielnik. "Mobility of nonlinear systems". Journal of Mathematical Physics **21**, 44–54 (1980).

[48] M Reginatto and M J W Hall. "Quantum-classical interactions and measurement: a consistent description using statistical ensembles on configuration space". Journal of Physics: Conference Series **174**, 012038 (2009).

[49] Lucien Hardy. "Probability theories with dynamic causal structure: a new framework for quantum gravity" (2005). arXiv:gr-qc/0509120.

[50] Giulio Chiribella, GM D'Ariano, Paolo Perinotti, and Benoit Valiron. "Beyond quantum computers" (2009). arXiv:0912.0195.

[51] Ognyan Oreshkov, Fabio Costa, and Časlav Brukner. "Quantum correlations with no causal order". Nature communications **3**, 1092 (2012).

[52] Eugene P Wigner. "Remarks on the mind-body question". In Philosophical reflections and syntheses. Pages 247–260. Springer (1995).

[53] Daniela Frauchiger and Renato Renner. "Quantum theory cannot consistently describe the use of itself". Nature communications **9**, 3711 (2018).

[54] Kok-Wei Bong, Aníbal Utreras-Alarcón, Farzad Ghafari, Yeong-Cherng Liang, Nora Tischler, Eric G. Cavalcanti, Geoff J. Pryde, and Howard M. Wiseman. "A strong no-go theorem on the wigner's friend paradox". Nature Physics **16**, 1199–1205 (2020).

[55] Eric G. Cavalcanti and Howard M. Wiseman. "Implications of local friendliness violation for quantum causality". Entropy**23** (2021).

[56] David Schmid, Yìlè Yīng, and Matthew Leifer. "A review and analysis of six extended wigner's friend arguments" (2023). arXiv:2308.16220.

[57] Yìlè Yīng, Marina Maciel Ansanelli, Andrea Di Biagio, Elie Wolfe, and Eric Gama Cavalcanti. "Relating wigner's friend scenarios to nonclassical causal compatibility, monogamy relations, and fine tuning" (2023). arXiv:2309.12987.

[58] GM D'Ariano, Franco Manessi, and Paolo Perinotti. "Determinism without causality". Physica Scripta **2014**, 014013 (2014).

[59] John H Selby, Maria E Stasinou, Stefano Gogioso, and Bob Coecke. "Time symmetry in quantum theories and beyond" (2022). arXiv:2209.07867.

[60] Matt Wilson, Giulio Chiribella, and Aleks Kissinger. "Quantum supermaps are characterized by locality" (2022). arXiv:2205.09844.

[61] Venkatesh Vilasini, Nuriya Nurgalieva, and Lídia del Rio. "Multi-agent paradoxes beyond quantum theory". New Journal of Physics **21**, 113028 (2019).

[62] Nick Ormrod, V Vilasini, and Jonathan Barrett. "Which theories have a measurement problem?" (2023). arXiv:2303.03353.

[63] Jonathan Barrett, Lucien Hardy, and Adrian Kent. "No signaling and quantum key distribution". Physical Review Letters **95**, 010503 (2005).

[64] Peter Janotta and Haye Hinrichsen. "Generalized probability theories: what determines the structure of quantum theory?". Journal of Physics A: Mathematical and Theoretical **47**, 323001 (2014).

[65] Martin Plávala. "General probabilistic theories: An introduction" (2021). arXiv:2103.07469.

[66] Giacomo Mauro D'Ariano, Paolo Perinotti, and Alessandro Tosini. "Information and disturbance in operational probabilistic theories" (2019). arXiv:1907.07043.

[67] Stephen D. Bartlett, Terry Rudolph, and Robert W. Spekkens. "Reference frames, superselection rules, and quantum information". Rev. Mod. Phys. **79**, 555–609 (2007).

[68] Mohammad Bahrami, André Großardt, Sandro Donadi, and Angelo Bassi. "The Schrödinger–Newton equation and its foundations". New Journal of Physics **16**, 115007 (2014).

[69] Heinz-Peter Breuer and F. Petruccione. "The theory of open quantum systems". Oxford University Press. Oxford ; New York (2002).

[70] E G Beltrametti and S Bugajski. “A classical extension of quantum mechanics”. Journal of Physics A: Mathematical and General **28**, 3329–3343 (1995).

[71] Daniel Carney and Jacob M. Taylor. “Strongly incoherent gravity” (2023). arXiv:2301.08378.

[72] Bogdan Mielnik. “Generalized quantum mechanics”. Comm. Math. Phys. **37**, 221–256 (1974).

[73] Asher Peres and Daniel Terno. “Hybrid classical-quantum dynamics”. Physical Review A **63**, 022101 (2001).

[74] John Selby and Bob Coecke. “Leaks: quantum, classical, intermediate and more”. Entropy **19**, 174 (2017).

[75] John H. Selby, Carlo Maria Scandolo, and Bob Coecke. “Reconstructing quantum theory from diagrammatic postulates”. Quantum **5**, 445 (2021).

[76] Bob Coecke, John Selby, and Sean Tull. “Two roads to classicality” (2017). arXiv:1701.07400.

A Semi-classical approximations

In general semi-classical approximations involve modelling some physical quantity by its expectation value. In some cases quantum fluctuations about the expectation value of the physical quantity are also included. Moreover these approximations are often accompanied by an assumption of locality, such as the Hartree-Fock approximation.

Example 5 (Mean field approximation)**.** In a mean field approximation an observable $\hat{O}_i$ is replaced by its mean value and some fluctuation about this value $\hat{O}_i \approx \langle O_i \rangle + \delta \hat{O}_i$. A Hamiltonian of the form $\hat{H} = \sum_i K_i + \sum_{\langle ij \rangle} \hat{O}_i \hat{O}_j$, where $\langle ij \rangle$ indicates nearest neighbours, in this approximation is $\hat{H} \approx \sum_{\langle ij \rangle} \left(\langle \hat{O}_j \rangle \hat{O}_i + \langle \hat{O}_i \rangle \hat{O}_j - \langle \hat{O}_i \rangle \langle \hat{O}_j \rangle \right)$. The time evolution generated by this Hamiltonian is a non-linear Schrodinger equation, as can be seen by the presence of state dependent terms in the Hamiltonian.

The term mean-field approximation is also used to describe the following approximation: first the Hartree-Fock approximation is made (multi-particle states are assumed to be product) and then one assumes that the Hamiltonian for a single system depends on the expectation value of some observable of the other particles (such as charge density). This leads to non-linear quantum master equations and non-linear Schrödinger equations [69, Section 3.7], such as the Gross-Pitaevskii equation for describing bosons.

The Hartree-Fock approximation and mean field limit can also be used in the case of quantum gravity. Assuming linearized quantum gravity (i.e., an effective quantum field-theoretic description of gravity in the weak-field limit) and making the assumptions described above one can derive a non-linear Schödinger equation (the Schödinger-Newton equation) as an *effective* description of the zeroth-order dynamics of N particles interacting with a quantum gravitational field in the aforementioned regime [68, Section 2.1].

B Schrödinger-Newton

The Schrödinger-Newton equation [68] appears in semi-classical gravity, where spacetime is described using the classical general relativistic framework and only matter degrees of freedom are quantized. Einstein’s equations describe the coupling between spacetime and matter in general relativity, and can be generalised to quantum matter by letting the expectation value of the stress energy tensor couple to the gravitational field:

$$R_{\mu\nu} + \frac{1}{2} g_{\mu\nu} R = \frac{8\pi G}{c^4} \langle \psi | \hat{T}_{\mu\nu} | \psi \rangle \, , \tag{7}$$

From this semi-classical Einstein equation one can derive the Schrödinger-Newton equation describing the evolution of the matter degrees of freedom [68]:

$$i\frac{d}{dt}\psi = \left[-\frac{\nabla^2}{2m} - Gm^2\int d^3r'\frac{|\psi(t,r')|^2}{|r-r'|} + V\right]\psi\,. \tag{8}$$

This setup seems to at first contradict our theorem. Namely matter is described by an irreducible quantum system, the interaction is reversible, there is back-reaction from the matter degrees of freedom onto the gravitational field and the gravitational field is classical.

This apparent contradiction is resolved once one observes that the non-linear dynamics described by the Schrödinger Newton equation allows one to carry out new non-quantum measurements on the matter system. These new measurements allow one to distinguish arbitrary pure states, entailing that the system is essentially classical. Therefore the Schrödinger-Newton model violates condition 1. of the theorem since it describes matter degrees of freedom as classical [47].

C Mielnik's generalized quantum mechanics

In his 1974 paper "Generalized quantum mechanics" [72] Mielnik studied modifications to the Schrödinger equation induced by 'subtle' interactions between a quantum system and a classical gravitational field. In the language of our theorem this corresponds to reversible interactions with back-reaction. In the light of our no-go theorem, Mielnik showed that imposing Conditions ii), iii) and that the gravitational field is classical led to a change in the nature of the quantum system. The new interactions allowed one to measure non-quadratic observables (i.e. non-quantum observables) which lead to the system being effectively classical, as in the case of the Schrödinger-Newton equation above. This led him to the following hypothesis: "Either gravity is quantum or the electron is classical".

Our theorem allows us to turn his hypothesis into a more precise statement: "Under the assumption of reversibility, either gravity is non-classical or the electron (i.e., matter) is a fundamentally super-selected (including classical) system".

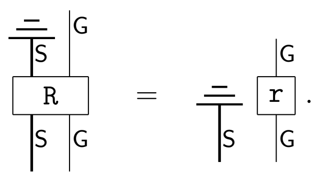
More specifically, if we assume that the position of the electron back-reacts on the gravitational field (as is predicted by Einstein's field equation) and that the interaction is reversible, then either gravity is non-classical or the position of the electron cannot be put into superpositions. A similar story can be told for the energy of the electron, or indeed, any other degree of freedom that we expect to back-react on the gravitational field. Note that this does not rule out the existence of non-classical degrees of freedom so long as they do not backreact on the gravitational field.

There exist other models of reversible interactions with back reaction between a classical system and a quantum system such as [73] where the quantum degree of freedom inherits classical features. Theorem 2 shows that this is inevitable (unless one allows that classical system to become non-classical).

D Constraining interactions between classical and non-classical systems

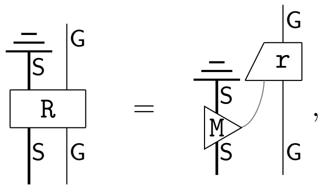
We now give the theorem constraining the possible interactions between non-classical systems and classical systems which we used to obtain Theorem 1 in the main text.

Theorem 2. There is no reversible interaction between a classical system, G, and fully non-classical GPT system, S with information flow from the GPT system to the classical system. That is, any such interaction R must satisfy the no-signalling condition



(9)

Moreover, in the case that the system S is reducible, then it is only the label of the sector which can influence the classical system in a reversible interaction. That is, it satisfies



(10)

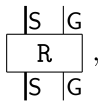
where M is the non-disturbing measurement for S which measures the which-sector information.

In the case of quantum theory this theorem shows that the only interactions between classical and non-superselected quantum systems where the quantum system influences the classical system are irreversible. Moreover, if we do have superselected quantum systems, where information other than the which-sector information to influences the classical system, then this too must be governed by an irreversible interaction.

Note that we expect the same proof to hold for the case of infinite dimensional classical systems G. On a formal level, what we need is to identify the relevant symmetric monoidal category for describing a generalised probabilistic theory coupled to infinite dimensional classical systems, at which point exactly the same proof that we give below should hold.

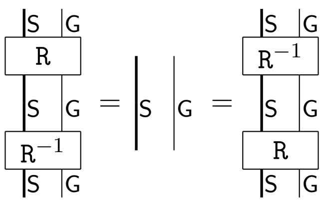
D.1 Proof of Theorem 2 for irreducible systems

Proof. Let us assume that we have a classical system G, which in our key application will correspond to a classical gravitational field, and a fully non-classical system S, which in our key application will correspond to the matter degree of freedom. Then, let us assume that these two systems are interacting via a reversible interaction R which we diagrammatically denote as:



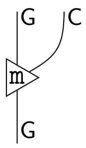
(11)

where reversibility means that there exists a physical transformation R^{-1} such that:



(12)

At least in principle, there exist classical measurements, m, of G which are non-disturbing[4]. We denote these as:

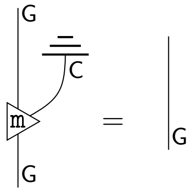


(13)

where C is some classical 'pointer' degree of freedom which records the measurement outcome. In the case that G is a finite dimensional classical system, then there is a single non-disturbing measurement which perfectly records the state of G into C, we, however, do not need to make this assumption here. The fact that this is a non-disturbing measurement is encoded in the fact that if we discard C the resulting transformation on F_g is simply the identity. Such a transformation is known as a *leak*

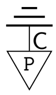
[4]This does not necessarily need to be a 'direct' measurement of the field, but could constitute a complicated 'indirect' procedure by which some test particle interacts with the field and then some property of the test particle is measured to give the result.

in the process-theory literature [74–76], in particular, one where the leaked system, C, is classical. Diagrammatically, these are characterised by the following equation:



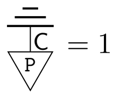
(14)

where  represents throwing away/discarding/ignoring the physical system, C. In quantum theory this is represented by the trace and in classical probability theory by marginalisation. In particular,



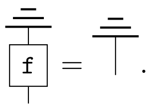
(15)

therefore represents the normalisation of the state P and hence any physical state must satisfy:



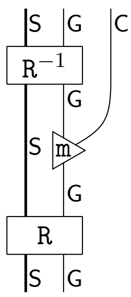
(16)

Similarly, any physical transformation must preserve this normalisation, hence we have that:



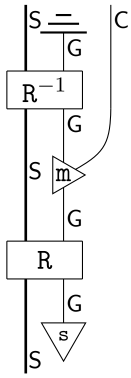
(17)

Now, consider the following diagram which we can construct out of these elements:



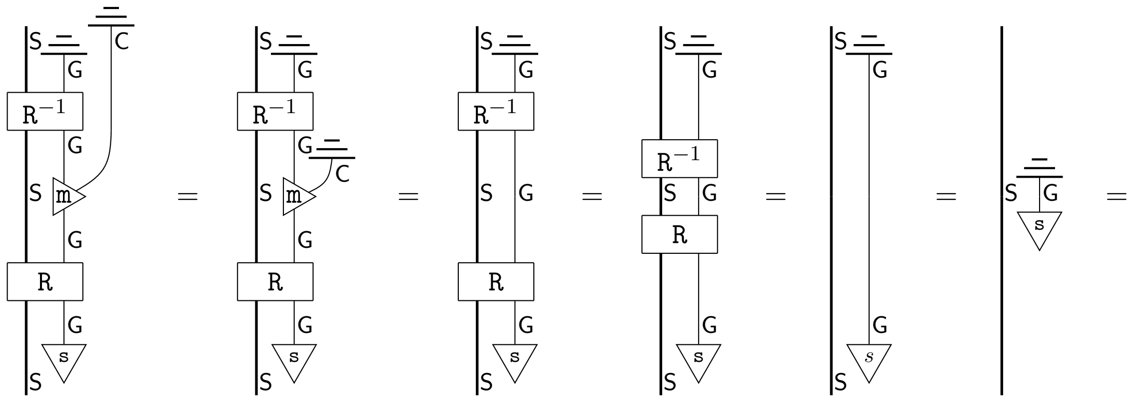
(18)

For some initial state of the field, s, and ignoring the final state of the field, we have the following:



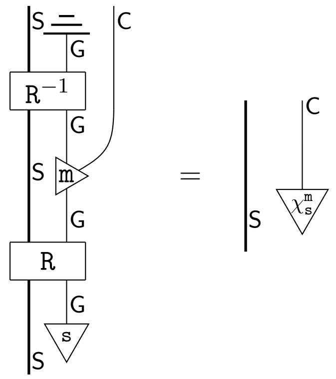
(19)

It is then simple to see that the above diagram describes a procedure that implements a non-disturbing measurement on the system S, i.e., it defines a leak for the system, as:

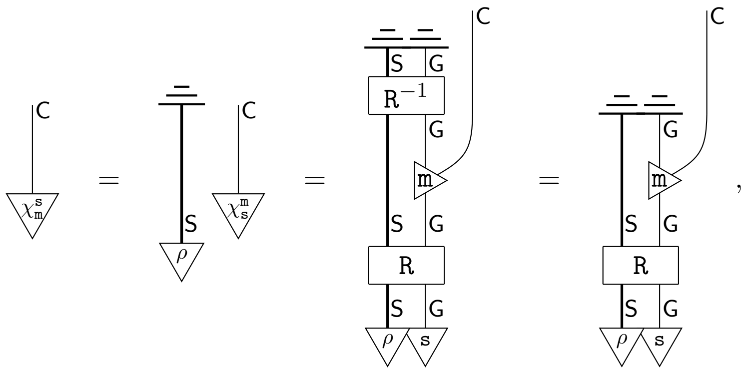
 (20)

where the second equality uses the fact that m is non-disturbing for the field, the fourth reversibility of R, and the sixth is the normalisation of the field state s.

Now, it is a well known result that such non-disturbing measurements for irreducible GPT systems are trivial [66], that is, that they factorise as the identity on the irreducible system and some fixed preparation of the classical pointer variable, that is:

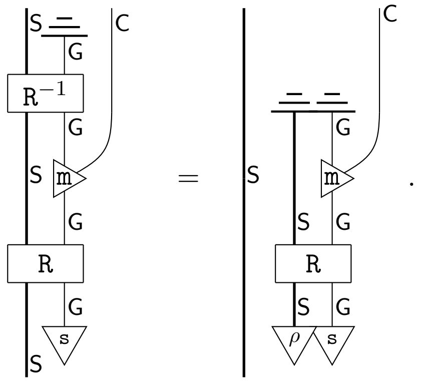
 (21)

where this state $\chi_{\mathtt{s}}^{\mathtt{m}}$ can depend on the initial state of the field s and the non-disturbing measurement of the field m. We can compute this state to be given by:

 (22)

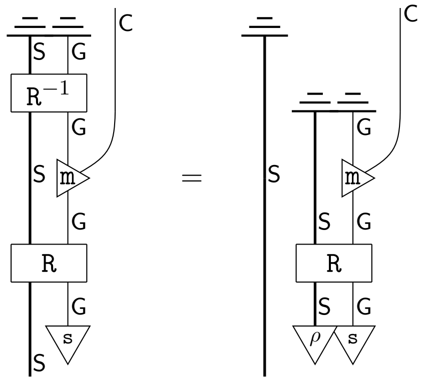
which is independent of the arbitrarily chosen initial state ρ. The first equality follows from normalisation of ρ, the second from Eq. (21), and the final one from the fact that $\mathtt{R}^{-1}$ is a physical transformation and so preserves normalisation.

Hence, we can rewrite eq. (21) as:



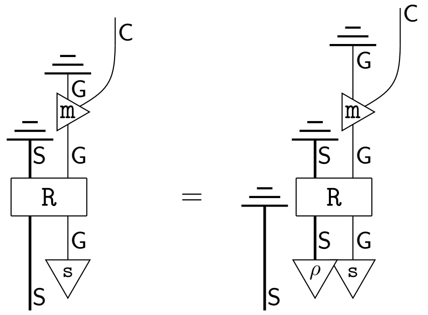
(23)

Now, if we trace out the irreducible system, S in Eq. (23), we obtain



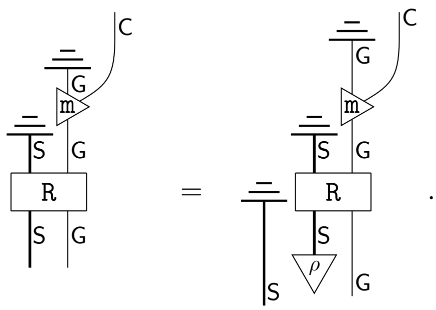
(24)

which, using the fact that $\mathtt{R}^{-1}$ is physical and hence preserves normalisation gives



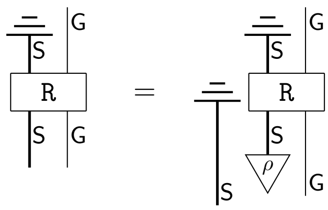
(25)

Now, this is true for all states s of the field, hence we have

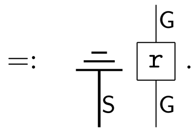


(26)

Similarly, this holds for all non-disturbing measurements of the field m, and as these suffice for tomography of the classical system G hence we can rewrite this as:



(27)



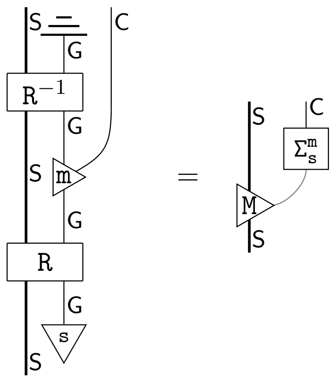
(28)

This is the definition of what it means for an interaction R to be no-signalling from S to G. In other words, the state of the system S before the interaction has no influence on the state of the classical gravitational field G after the interaction. ☐

D.2 Proof of Theorem 2 for reducible systems

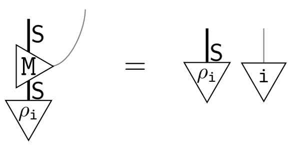
Up to Eq. 21 the proof is identical, as that is the first point where we have invoked irreducibility of S, the rest of the proof then proceeds as follows:

Proof. Now, it is a well known result that such non-disturbing measurements for GPT systems reveal only which-sector information [66], that is, that they can be written as:



(29)

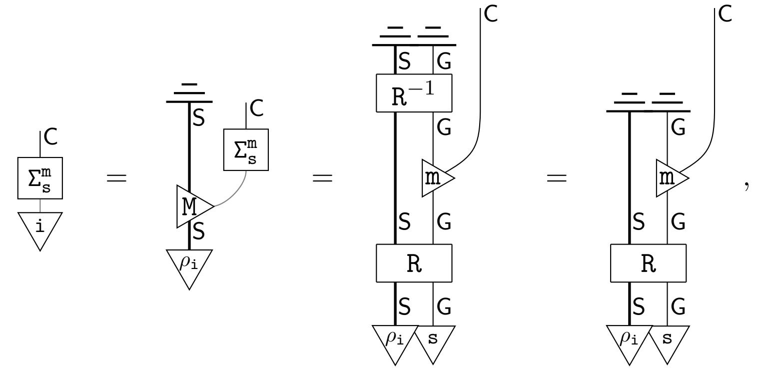
where M is the non-disturbing measurement that reveals the which-sector information, that is, it satisfies:



(30)

for all i and $\rho_i \in \Omega_i$. The process Σ^{m}_{s} is a stochastic map (which can depend on the initial state of the field s and the non-disturbing measurement of the field m) processing the which-sector information.

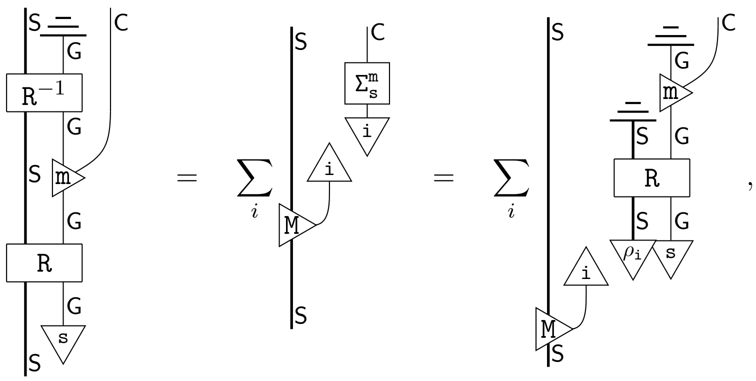
We can compute this stochastic map by looking at its action on point distributions as:



(31)

where $\rho_{\mathtt{i}}$ is an arbitrarily chosen state in sector Ω_i. The first equality follows from the definition of $\mathtt{M}$ together with the normalisation of $\rho_{\mathtt{i}}$, the second from Eq. (29), and the final one from the fact that $\mathtt{R}^{-1}$ is a physical transformation and so preserves normalisation.

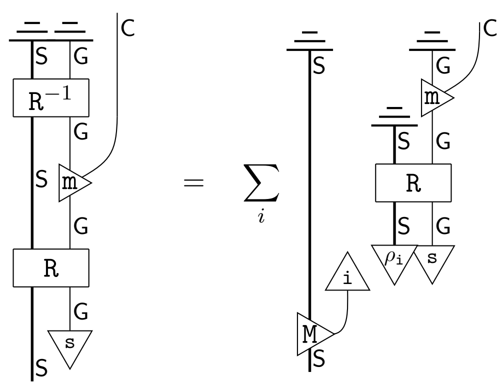
Hence, we can rewrite eq. (29) as:



(32)

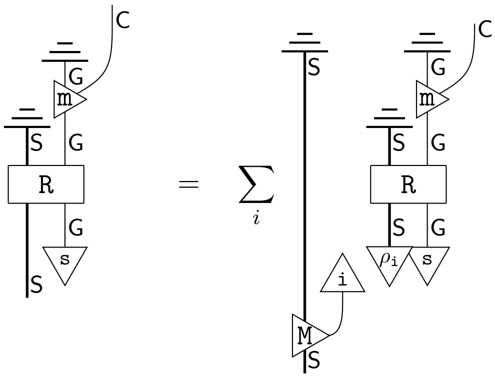
where in the first step we have taken Eq. (29) and introduced a resolution of the identity on the RHS, and in the second we have used the above characterisation of $\Sigma^{\mathtt{m}}_{\mathtt{s}}$.

Now, if we trace out the system, S in Eq. (32), we obtain



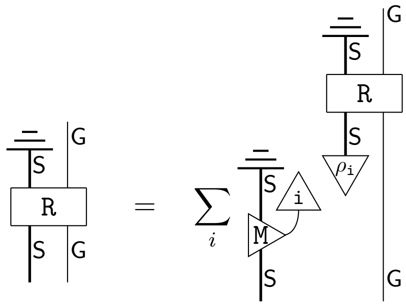
(33)

which, using the fact that $\mathtt{R}^{-1}$ is physical and hence preserves normalisation gives

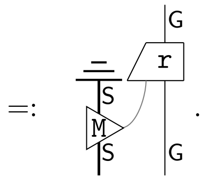


(34)

Now, this is true for all states $\mathtt{s}$ and measurements $\mathtt{m}$ of the field, hence we have:



(35)



(36)

This means it is only the variation in the which-sector information of S which signals to G. □

E Algebraic version of the proof for classical-quantum systems

E.1 Algebraic proof for irreducible systems

For the reader who is not familiar with diagrammatic notation and GPTs we offer an algebraic version of the proof of Theorem 2 for irreducible systems in the case where the non-classical system is quantum.

This proof broadly follows the diagrammatic proof of Section D.1 and we refer to the diagrammatic version of equations so that the interested reader can translate between the diagrammatic approach and the algebraic approach.

Let S be a non-super-selected quantum system $\mathbb{C}^d$ and G a finite n-dimensional classical system. We represent G using diagonal density operators on $\mathbb{C}^n$.

A general classical quantum state is of the following form:

$$\tilde{\rho}_{\mathsf{GS}} = \sum_x p(x)\, |x\rangle\langle x|_{\mathsf{G}} \otimes \rho(x)_{\mathsf{S}}, \tag{37}$$

where $p(x)$ is some probability distribution and $\rho(x)_{\mathsf{S}}$ a density operator. The interaction $R : \mathsf{SG} \to \mathsf{SG}$ is reversible hence there exists a transformation $R^{-1} : \mathsf{SG} \to \mathsf{SG}$ such that

$$R^{-1}(R^{-1}(\cdot)) = \mathbb{I}_{\mathsf{GS}}, \tag{38}$$

with $\mathbb{I}_{\mathsf{GS}}$ the identity super-operator on the classical quantum states.

R is assumed to be signalling from $\mathsf{S} \to \mathsf{G}$ hence for some input states ρ_{S} and $\sum_x p(x) |x\rangle\langle x|_{\mathsf{G}}$ the output state is:

$$R\left(\sum_x p(x) |x\rangle\langle x|_{\mathsf{G}} \otimes \rho_{\mathsf{S}}\right) = \sum_x q(x) |x\rangle\langle x| \otimes \sigma_{\mathsf{S}}(x), \tag{39}$$

where the reduced state $\sum_x q(x) |x\rangle\langle x|_{\mathsf{C}}$ depends on the input state ρ_{S}.

The leak of Equation (13) in the case where the measurement is the canonical measurement, i.e. the measurement in the $|x\rangle\langle x|_{\mathsf{G}}$ basis is given by the following map:

$$m : \mathsf{G} \to \mathsf{CG} \tag{40}$$

$$m :: \sum_x p(x) |x\rangle\langle x|_{\mathsf{G}} \mapsto \sum_x p(x) |x\rangle\langle x|_{\mathsf{G}} \otimes |x\rangle\langle x|_{\mathsf{C}}\,, \tag{41}$$

where C is a classical system of the same dimension as G. This operation can be understood as a sort of classical copy operation since the reduced state on C after the transformation m is equal to the initial state of G. We stress that the diagrammatic proof is more general and applies to any leak, not the one corresponding to the canonical measurement only.

We observe that tracing out C after m is just the identity on G:

$$\mathrm{Tr}_{\mathsf{C}}\left(m\left(\sum_x p(x) |x\rangle\langle x|_{\mathsf{G}}\right)\right) = \mathrm{Tr}_{\mathsf{C}}\left(\sum_x p(x) |x\rangle\langle x|_{\mathsf{G}} \otimes |x\rangle\langle x|_{\mathsf{C}}\right) = \sum_x p(x) |x\rangle\langle x|_{\mathsf{G}}\,, \tag{42}$$

showing that this obeys the condition of Equation (14). The diagram of Equation (18) is translated to:

$$E : \mathsf{SG} \to \mathsf{SGC} \tag{43}$$

$$E :: \rho_{\mathsf{GS}} \mapsto (R^{-1} \otimes \mathbb{I}_{\mathsf{C}})\,(m(R(\rho_{\mathsf{GS}})))\,. \tag{44}$$

The interaction of Equation (19) is obtained from E by fixing an initial state $s_{\mathsf{G}} = \sum_x p(x) |x\rangle\langle x|_{\mathsf{G}}$ of G and tracing out the state of G after E, this defines a new map:

$$E_{s_{\mathsf{G}}} : \mathsf{S} \to \mathsf{SC} \tag{45}$$

$$E_{s_{\mathsf{G}}} :: \rho_{\mathsf{S}} \mapsto \mathrm{Tr}_{\mathsf{G}}\left(E\left(\sum_x p(x) |x\rangle\langle x| \otimes \rho_{\mathsf{S}}\right)\right). \tag{46}$$

We see that this map is a non-demolition measurement of S since it takes S to some joint state on SC where C is the classical pointer system. Following Equation (20) we can show that this measurement is non-disturbing:

$$\mathrm{Tr}_{\mathsf{C}}(E_{s_{\mathsf{G}}}(\cdot)) : \mathsf{S} \to \mathsf{S} \tag{47}$$

$$\mathrm{Tr}_{\mathsf{C}}(E_{s_{\mathsf{G}}}(\cdot)) :: \rho_{\mathsf{S}} \mapsto \mathrm{Tr}_{\mathsf{GC}}\left(E\left(\sum_x p(x)\,|x\rangle\langle x| \otimes \rho_{\mathsf{S}}\right)\right) \tag{48}$$

$$= \mathrm{Tr}_{\mathsf{GC}}\left((R^{-1} \otimes \mathbb{I}_{\mathsf{C}})\left(m\left(U_{\mathsf{GS}}\left(\sum_x p(x)\,|x\rangle\langle x| \otimes \rho_{\mathsf{S}}\right)U_{\mathsf{GS}}^{\dagger}\right)\right)\right) \tag{49}$$

$$= \mathrm{Tr}_{\mathsf{GC}}\left((R^{-1} \otimes \mathbb{I}_{\mathsf{C}})\left(m\left(\sum_x q(x)\,|x\rangle\langle x| \otimes \sigma_{\mathsf{S}}(x)\right)\right)\right) \tag{50}$$

$$= \mathrm{Tr}_{\mathsf{GC}}\left((R^{-1} \otimes \mathbb{I}_{\mathsf{C}})\left(\sum_x q(x)\,|x\rangle\langle x| \otimes \sigma_{\mathsf{S}}(x) \otimes |x\rangle\langle x|_{\mathsf{C}}\right)\right) \tag{51}$$

$$= \mathrm{Tr}_{\mathsf{G}}\left(R^{-1}\left(\mathrm{Tr}_{\mathsf{C}}\left(\sum_x q(x)\,|x\rangle\langle x| \otimes \sigma_{\mathsf{S}}(x) \otimes |x\rangle\langle x|_{\mathsf{C}}\right)\right)\right) \tag{52}$$

$$= \mathrm{Tr}_{\mathsf{G}}\left(R^{-1}\left(\sum_x q(x)\,|x\rangle\langle x| \otimes \sigma_{\mathsf{S}}(x)\right)\right) \tag{53}$$

$$= \mathrm{Tr}_{\mathsf{G}}\left(\sum_x p(x)\,|x\rangle\langle x|_{\mathsf{G}} \otimes \rho_{\mathsf{S}}\right) \tag{54}$$

$$= \rho_{\mathsf{S}}, \tag{55}$$

where we have used the fact that $\mathrm{Tr}_{\mathsf{GC}}\left((R^{-1} \otimes \mathbb{I}_{\mathsf{C}})(\cdot)\right) = \mathrm{Tr}_{\mathsf{G}}\left((R^{-1}(\mathrm{Tr}_{\mathsf{C}}(\cdot))\right)$. This is the first step in Equation (20).

Let us compute the state of the pointer system C after the interaction by tracing out the state of GS:

$$\mathrm{Tr}_{\mathsf{GS}}(E_{s_{\mathsf{G}}}(\cdot)) : \mathsf{S} \to \mathsf{C} \tag{56}$$

$$\mathrm{Tr}_{\mathsf{GS}}(E_{s_{\mathsf{G}}}(\cdot)) :: \rho_{\mathsf{S}} \mapsto \mathrm{Tr}_{\mathsf{GS}}\left(E\left(\sum_x p(x)\,|x\rangle\langle x| \otimes \rho_{\mathsf{S}}\right)\right) \tag{57}$$

$$= \mathrm{Tr}_{\mathsf{GS}}\left((R^{-1} \otimes \mathbb{I}_{\mathsf{C}})\left(m\left(U_{\mathsf{GS}}\left(\sum_x p(x)\,|x\rangle\langle x| \otimes \rho_{\mathsf{S}}\right)U_{\mathsf{GS}}^{\dagger}\right)\right)\right), \tag{58}$$

and we use the fact that R^{-1} preserves normalisation, so that $\mathrm{Tr}_{\mathsf{SG}}(R^{-1}(\rho_{\mathsf{SG}})) = \mathrm{Tr}_{\mathsf{SG}}(\rho_{\mathsf{SG}})$ (the final step of Equation (22)), then

$$\mathrm{Tr}_{\mathsf{GS}}(E_{s_{\mathsf{G}}}(\rho_{\mathsf{S}})) = \mathrm{Tr}_{\mathsf{GS}}\left(m\left(U_{\mathsf{GS}}\left(\sum_x p(x)\,|x\rangle\langle x| \otimes \rho_{\mathsf{S}}\right)U_{\mathsf{GS}}^{\dagger}\right)\right) \tag{59}$$

$$= \mathrm{Tr}_{\mathsf{GS}}\left(m\left(\sum_x q(x)\,|x\rangle\langle x| \otimes \sigma_{\mathsf{S}}(x)\right)\right) \tag{60}$$

$$= \mathrm{Tr}_{\mathsf{GS}}\left(\sum_x q(x)\,|x\rangle\langle x| \otimes \sigma_{\mathsf{S}}(x) \otimes |x\rangle\langle x|_{\mathsf{C}}\right) \tag{61}$$

$$= \sum_x q(x)\,|x\rangle\langle x|_{\mathsf{C}}\,. \tag{62}$$

Observe that since we assume that there is information flow from $\mathsf{S} \to \mathsf{G}$ during the interaction R, this implies that for at least one input state ρ_{S} and $\sum_x p(x)\,|x\rangle\langle x|_{\mathsf{G}}$ the state $\sum_x q(x)\,|x\rangle\langle x|_{\mathsf{C}}$ depends non-trivially on ρ_{S}.

Thus $\mathrm{Tr}_{\mathsf{GS}}(E_{s_{\mathsf{G}}}(\cdot)) : \mathsf{S} \to \mathsf{C}$ is a non-trivial measurement of S.

This is a contradiction since we have created an interaction $E_{s_{\mathsf{G}}}(\cdot) : \mathsf{S} \to \mathsf{SC}$ which is non-trivial non-disturbing measurement for an irreducible system.

E.2 The role of the no-signalling assumption

Observe that

$$(R^{-1} \otimes \mathbb{I}_{\mathsf{C}}) \left(\sum_x q(x) \, |x\rangle\langle x| \otimes \sigma_{\mathsf{S}}(x) \otimes |x\rangle\langle x|_{\mathsf{C}} \right) = \sum_x p(x) \, |x\rangle\langle x|_{\mathsf{G}} \otimes \rho_{\mathsf{S}} \otimes |x\rangle\langle x|_{\mathsf{C}} \, . \tag{63}$$

However the state of C before the interaction was $\sum_x p(x) \, |x\rangle\langle x|_{\mathsf{C}}$ whereas after the interaction it is $\sum_x q(x) \, |x\rangle\langle x|_{\mathsf{C}}$. This is in general inconsistent with the fact that the action of $R^{-1} \otimes \mathbb{I}_{\mathsf{C}}$ should leave C unchanged.

The assumption that the action of $R^{-1} \otimes \mathbb{I}_{\mathsf{C}}$ should leave C unchanged is sometimes known as no-signalling, since it says that an operation on SG should not affect C. It is met by quantum theory and is an assumption of the GPT framework more generally.

If we impose no-signalling then the interaction is only consistent when $p(x) = q(x)$ and hence

$$R :: \sum_x p(x) \, |x\rangle\langle x| \otimes \rho_{\mathsf{S}} \mapsto \sum_x p(x) \, |x\rangle\langle x| \otimes \sigma_{\mathsf{S}}(x), \tag{64}$$

i.e. there is no back-reaction from S to G. This is consistent with the main theorem.