Thermodynamics and the structure of quantum theory

2017 New J. Phys. 19 043025

(http://iopscience.iop.org/1367-2630/19/4/043025)

IOP Publishing New J. Phys. 19 (2017) 043025 https://doi.org/10.1088/1367-2630/aa68ef

New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics

Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

PAPER

Thermodynamics and the structure of quantum theory

Marius Krumm[1,2,3,8], Howard Barnum[4,5], Jonathan Barrett[6] and Markus P Müller[1,2,3,7]

[1] Department of Applied Mathematics, University of Western Ontario, London, ON N6A 5BY, Canada
[2] Department of Theoretical Physics, University of Heidelberg, Heidelberg, Germany
[3] Perimeter Institute for Theoretical Physics, Waterloo, ON N2L 2Y5, Canada
[4] Department of Physics and Astronomy, University of New Mexico, Albuquerque, NM, United States of America
[5] Institute for Theoretical Physics and Riemann Center for Geometry and Physics, Leibniz University Hannover, Appelstraße 2, D-30163 Hannover, Germany
[6] Department of Computer Science, University of Oxford, Oxford, United Kingdom
[7] Department of Philosophy, University of Western Ontario, London, ON N6A 5BY, Canada
[8] Author to whom any correspondence should be addressed.

E-mail: mkrumm@uwo.ca

CrossMark

OPEN ACCESS

RECEIVED
20 November 2016

REVISED
5 February 2017

ACCEPTED FOR PUBLICATION
24 March 2017

PUBLISHED
19 April 2017

Keywords: general probabilistic theories, quantum thermodynamics, higher-order interference, von Neumann's thought experiment, second law of thermodynamics, thermodynamic entropy

Abstract

Despite its enormous empirical success, the formalism of quantum theory still raises fundamental questions: why is nature described in terms of complex Hilbert spaces, and what modifications of it could we reasonably expect to find in some regimes of physics? Here we address these questions by studying how compatibility with thermodynamics constrains the structure of quantum theory. We employ two postulates that any probabilistic theory with reasonable thermodynamic behaviour should arguably satisfy. In the framework of generalised probabilistic theories, we show that these postulates already imply important aspects of quantum theory, like self-duality and analogues of projective measurements, subspaces and eigenvalues. However, they may still admit a class of theories beyond quantum mechanics. Using a thought experiment by von Neumann, we show that these theories admit a consistent thermodynamic notion of entropy, and prove that the second law holds for projective measurements and mixing procedures. Furthermore, we study additional entropy-like quantities based on measurement probabilities and convex decomposition probabilities, and uncover a relation between one of these quantities and Sorkin's notion of higher-order interference.

1. Introduction

Quantum mechanics has existed for about 100 years now, but despite its enormous success in experiment and application, the meaning and origin of its counterintuitive formalism is still widely considered to be difficult to grasp. Many attempts to put quantum mechanics on a more intuitive footing have been made over the decades, which includes the development of a variety of interpretations of quantum physics (such as the many-worlds interpretation [1], Bohmian mechanics [2], QBism [3], and many others [4]), and a thorough analysis of its departure from classical physics (as in Bell's theorem [5] or in careful definitions of notions of contextuality [6]). In more recent years, researchers, mostly coming from and inspired by the field of quantum information processing (early examples include [21, 22, 51]), have taken as a starting point the set of all probabilistic theories. Quantum theory is one of them and can be uniquely determined by specifying some of its characteristic properties [53] (as in e.g. [19, 43, 51, 54, 55, 57–61]).

While the origins of this framework date back at least to the 1960s [15, 16, 18], it was the development of quantum information theory with its emphasis on simple operational setups that led to a new wave of interest in 'generalised probabilistic theories' (GPTs) [51, 52]. This framework turned out to be very fruitful for fundamental investigations of quantum theory's information-theoretic and operational properties. For example, GPTs make it possible to contrast quantum information theory with other possible theories of information processing, and in this way to gain a deeper understanding of its characteristic properties in terms of computation or communication.

In a complementary approach, there has been a wave of attempts to find simple physical principles that single out quantum correlations from the set of all non-signalling correlations in the device-independent formalism [70]. These include non-trivial communication complexity [71], macroscopic locality [72], or information causality [73]. However, none of these principles so far turns out to yield the set of quantum correlations exactly. This led to the discovery of 'almost quantum correlations' [75] which are more general than those allowed by quantum theory, but satisfy all the aforementioned principles. Almost quantum correlations seem to appear naturally in the context of quantum gravity [77].

A relation to other fields of physics can also be drawn from information causality, which can be understood as the requirement that a notion of entropy [66–69] exists which has some natural properties like the data-processing inequality [74]. These emergent connections to entropy and quantum gravity are particularly interesting since they point to an area of physics where modifications of quantum theory are well-motivated: Jacobson's results [78] and holographic duality [79] relate thermodynamics, entanglement, and (quantum) gravity, and modifying quantum theory has been discussed as a means to overcome apparent paradoxes in black-hole physics [80].

While GPTs provide a way to generalise quantum theory and to study more general correlations and physical theories, they still leave open the question as to which principles should guide us in applying the GPT formalism for this purpose. The considerations above suggest taking, as a guideline for such modifications, the principle that they support a well-behaved notion of thermodynamics. As A Einstein [32] put it,

'A theory is the more impressive the greater the simplicity of its premises, the more different kinds of things it relates, and the more extended its area of applicability. Therefore the deep impression that classical thermodynamics made upon me. It is the only physical theory of universal content which I am convinced will never be overthrown, within the framework of applicability of its basic concepts.'

Along similar lines, A Eddington [33] argued that *'The law that entropy always increases holds, I think, the supreme position among the laws of Nature. If someone points out to you that your pet theory of the Universe is in disagreement with Maxwell's equations—then so much the worse for Maxwell's equations. If it is found to be contradicted by observation—well, these experimentalists do bungle things sometimes. But if your theory is found to be against the second law of thermodynamics I can give you no hope; there is nothing for it but to collapse in deepest humiliation.'*

Here we take this point of view seriously. We investigate what kinds of probabilistic theories, including but not limited to quantum theory, could peacefully coexist with thermodynamics. We present two postulates that formalise important physical properties which can be expected to hold in any such theory. On the one hand, these two postulates allow for a class of theories more general than quantum or classical theory, which thus describes potential alternative physics consistent with important parts of thermodynamics as we know it. Indeed, by considering a thought experiment originally conceived by von Neumann, we show that these theories all give rise to a unique, consistent form of thermodynamical entropy. Furthermore, we show that this entropy satisfies several other important properties, including two instances of the second law. On the other hand, we show that these postulates already imply many structural properties which are also present in quantum theory, for example self-duality and the existence of analogues of projective measurements, observables, eigenvalues and eigenspaces.

In summary, our analysis shows that important structural aspects of quantum and classical theory are already implied by these aspects of thermodynamics, but on the other hand it suggests that there is still some 'elbow room' for modification within these limits dictated by thermodynamics.

Thermodynamics in GPTs has been considered in some earlier works. In [35, 36], the authors introduced a notion of (Rényi-2-)entanglement entropy, and studied the phenomenon of thermalisation by entanglement [37–39] and the black-hole information problem (in particular the Page curve [40]) in generalisations of quantum theory. Hänggi and Wehner [46] have related the uncertainty principle to the second law in the framework of GPTs. Chiribella and Scandolo ([45, 47], see also [48]) have considered the notion of diagonalization and majorization in general theories, leading to a resource-theoretic approach to thermodynamics in GPTs. There are various connections between their results and ours, but there are essential differences. In particular, they assume the purification postulate (which is arguably a strong assumption that in particular excludes classical thermodynamics), whereas we are not making any assumption on composition of systems whatsoever, and in this sense work in a more general framework. Furthermore, while Chiribella and Scandolo take a resource-theoretic approach motivated by quantum information theory, our analysis relies on a more traditional thermodynamical thought experiment (namely von Neumann's). We presented results related to some of those in the present paper in the conference proceedings [31]; here we use different assumptions and obtain additional results.

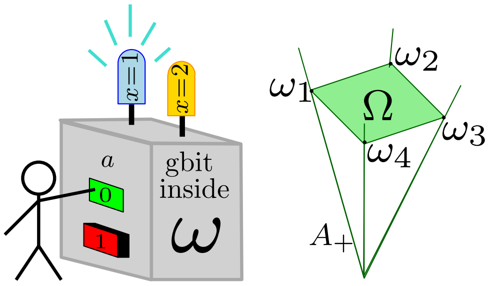

Figure 1. An example state space, A, modelling a so-called 'gbit' [52] which is often used to describe one half of a PR-box. The operational setup is depicted on the left, and the mathematical formulation is sketched on the right. An agent ('Alice') holds a black box ω into which she can input one bit, $a \in \{0, 1\}$, and obtains one output, $x \in \{1, 2\}$. The box is described by a conditional probability $p(x|a)$. In the GPT framework, ω becomes an actual state, i.e. an element of some state space Ω. Concretely, $\omega = (1, p(1|0), p(1|1)) \in \mathbb{R}^3$, where the first entry 1 is used to describe the normalisation, $p(1|0) + p(2|0) = p(1|1) + p(2|1)$. In this case, all probabilities are allowed by definition, so that the state space Ω becomes the square, i.e. the points $(1, s, t)$ with $0 \leqslant s, t \leqslant 1$. Alice's input a is interpreted as a 'choice of measurement', and the two measurements are $e_{x=1}^{(a=0)}, e_{x=2}^{(a=0)}$ resp. $e_{x=1}^{(a=1)}, e_{x=2}^{(a=1)}$ such that $\sum_{x=1}^{2} e_x^{(a)}(\omega) = 1$ for all states $\omega \in \Omega$. If we describe effects by vectors by using the standard inner product, we have, for example, $e_{x=1}^{(a=0)} = (0, 1, 0)$, since $e_{x=1}^{(a=0)}(\omega) = P(1|0) = (0, 1, 0) \cdot \omega$. There are four pure states, labelled $\omega_1, \ldots, \omega_4$. Every pure state ω_i is perfectly distinguishable from every other pure state ω_j for $j \neq i$, but no more than two of them are jointly distinguishable in a single measurement. More generally, every state on one side of the square is perfectly distinguishable from every state on the opposite side. The unit effect is $u_A = (1, 0, 0)$.

Our paper is organised as follows. We start with an overview of the framework of GPTs. Then we present von Neumann's thought experiment on thermodynamic entropy, and a modification of it due to Petz [42]. Although it relies on very mild assumptions, it already rules out all theories that admit a state space known as the *gbit* or *squit* (a square-shaped state space that can be used to describe one of the two local subsytems of a composite system known as the PR-box [83], exhibiting stronger-than-quantum correlations). Then we present our two postulates, and show that they imply many structural features of quantum theory. We show that theories that satisfy both postulates behave consistently in von Neumann's thought experiment and admit a notion of thermodynamic entropy which satisfies versions of the second law.

Because entropies are an important bridge between information theory and thermodynamics, in the final section we investigate the consequences of our postulates for generalisations of quantities of known significance in quantum thermodynamics [30], defined by applying Rényi entropies to probabilities in convex decompositions of a state, or of measurements made on a state. In particular, we show a relation between max-entropy and Sorkin's notion of higher-order interference [76]: equality of the preparation and measurement based max-entropies implies the absence of higher-order interference. Most proofs are deferred to the appendix. Several results of this paper have been announced in the Master Thesis of one of the authors [34].

2. The mathematical framework

Our results are obtained in the framework of GPTs [51, 52, 55, 85, 88]. The goal of this framework is to capture all probabilistic theories, i.e. all theories that use states to make predictions for probabilities of measurement outcomes. Although the framework is based on very weak and natural assumptions, we can only provide a short introduction of the main notions and results here. For more detailed explanations of the framework, see e.g. [34, 51, 52, 55, 86, 87]. The framework contains quantum theory and also the application of probability theory to classical physics, often referred to as classical probability theory, as special cases. It also contains theories which differ substantially from classical or quantum probability theory, for example boxworld [52], which allows superstrong nonlocality, and theories that allow higher-order interference [76].

A central notion is that of the state and the set of states, the state space Ω_A. A state contains all information necessary to calculate all probabilities for all outcomes of all possible measurements. One possible and convenient representation would be to simply list the probabilities of a set of 'fiducial' measurement outcomes which is sufficient to calculate all outcome probabilities for all measurements [51, 52]. An example is given in figure 1.

It is possible to create statistical mixtures of states: let us assume a black box device randomly prepares a state ω_1 with probability p_1 and a state ω_2 with probability p_2. In agreement with the representation of states as lists of probabilities and the law of total probability, the appropriate state to describe the resulting measurement statistics is $\omega = p_1\omega_1 + p_2\omega_2$. This means that the state space Ω_A is convex and is embedded into a real vector space A (to be described below). Due to the interpretation of states as lists of probabilities (which are between 0 and 1) we demand that Ω_A is bounded. Any state that cannot be written as a convex decomposition of other states is called a pure state. As pure states cannot be interpreted as statistical mixtures of other states, they are also called states of maximal knowledge. Furthermore, there is no physical distinction between states that can be prepared exactly, and states that can be prepared to arbitrary accuracy. Thus, we also assume that Ω_A is topologically closed. In order to not obscure the physics by the mathematical technicalities introduced by infinite dimensions, we will assume that A is finite-dimensional. Thus Ω_A is compact. Consequently, every state can be obtained as a statistical mixture of finitely many pure states [89].

Furthermore, it turns out to be convenient to introduce unnormalised states ω, defined as the non-negative multiples of normalised states. They form a closed convex cone $A_+ := \mathbb{R}_{\geqslant 0} \cdot \Omega_A$. For simplicity of description, we choose the vector space containing the cone of states to be of minimal dimension, i.e. $\mathrm{span}(A_+) = A$.

We introduce the normalisation functional $u_A : A \to \mathbb{R}$ which attains the value one on all normalised states, i.e. $u_A(\omega) = 1$ for all $\omega \in \Omega_A$. It is linear, non-negative on the whole cone, zero only for the origin, and $\omega \in A_+$ is an element of Ω_A if and only if $u_A(\omega) = 1$. The normalisation $u_A(\omega)$ can be interpreted as the probability of success of the preparation procedure. For states with $u_A(\omega) < 1$, the preparation succeeds with probability $u_A(\omega)$. The states with normalisation >1 do not have a physical interpretation, but adding them allows us to take full advantage of the notion of cones from convex geometry.

Effects are functionals that map (sub)normalised states to probabilities, i.e. into $[0, 1]$. To each measurement outcome we assign an effect that calculates the outcome probability for any state. Effects have to be linear for consistency with the statistical mixture interpretation of convex combinations of states. A *measurement* (with n outcomes) is a collection of effects $e_1, \ldots, e_n$ such that $e_1 + \ldots + e_n = u_A$. Its interpretation is that performing the measurement on some state $\omega \in \Omega_A$ yields outcome i with probability $e_i(\omega)$.

A set of states $\omega_1, \ldots, \omega_n$ is called *perfectly distinguishable* if there exists a measurement $e_1, \ldots, e_n$ such that $e_i(\omega_j) = \delta_{ij}$, that is, 1 if $i = j$ and 0 otherwise. A collection of n perfectly distinguishable pure states is called an *n-frame*, and a frame is called *maximal* if it has the maximal number n of elements possible in the given state space. In quantum theory, for example, the maximal frames are exactly the orthonormal bases of Hilbert space. In more detail, a frame on an N-dimensional quantum system is given by $\omega_1 = |\psi_1\rangle\langle\psi_1|, \ldots, |\psi_N\rangle\langle\psi_N|$, where $|\psi_1\rangle, \ldots, |\psi_N\rangle$ are orthonormal basis vectors.

Transformations are maps $T : A \to A$ that map states to states, i.e. $T(A_+) \subseteq A_+$. Similarly as effects, they also have to be linear in order to preserve statistical mixtures. They cannot increase the total probability, but are allowed to decrease it (as is the case, for example, for a filter), thus $u_A \circ T(\omega) \leqslant u_A(\omega)$ for all $\omega \in A_+$.

Instruments[9] [84] are collections of transformations T_j such that $\sum_j u_A \circ T_j = u_A$. If an instrument is applied to a state ω, one obtains outcome j (and post-measurement state $T_j(\omega)/p_j$) with probability $p_j := u_A(T_j(\omega))$. Each instrument corresponds to a measurement given by the effects $u_A \circ T_j$. We will say it 'induces' this measurement.

The framework of GPTs does not assume *a priori* that all mathematically well-defined states, transformations and measurements can actually be physically implemented. Here, we will assume that a measurement constructed from physically allowed effects is also physically allowed. Moreover, we assume that the set of allowed effects has the same dimension as A_+, because otherwise there would be distinct states that could not be distinguished by any measurement.

3. von Neumann's thought experiment

The following thought experiment has been applied by von Neumann [41] to find a notion of thermodynamic entropy for quantum states ρ. The result turns out to equal von Neumann entropy, $H(\rho) = -\mathrm{tr}(\rho \log \rho)$. We apply the thought experiment to a wider class of probabilistic theories.

We adopt the physical picture used by von Neumann [41] to describe the thought experiment[10]; we will comment on some idealisations used in this model at the end of this section. We consider a GPT ensemble $[S_1, \ldots, S_N]$, where S_i denotes the ith physical system, and N_j of the systems are in state ω_j, where $j = 1, \ldots, n$ and $\sum_j N_j = N$. This ensemble is described by the state $\omega = \sum_{j=1}^{n} p_j \omega_j$, where $p_j = N_j/N$, which describes the effective state of a system that is drawn uniformly at random from the ensemble.

[9] Some authors have recently begun referring to instruments as *operations*, but long-standing convention in quantum information theory (including [50]) uses the term 'operation' for the quantum case of what we are calling transformations (which are completely positive maps). Also, Davies and Lewis [84] define instrument more generally, to allow for continuously-indexed transformations, where we only consider finite collections T_j.

[10] Our thought experiment is identical to von Neumann's, up to two differences: first, we translate all quantum notions to more general GPT notions; second, while von Neumann implements the transition from (5) to (6) in figure 2 via sequences of projections, we implement this transition directly via reversible transformations.

We introduce N small, indistinguishable, hollow boxes[11], and we put each ensemble system S_j into one of the boxes such that the system is completely isolated from the outside. Furthermore, we assume that the boxes form an ideal gas, which will allow us to use the ideal gas laws in the following derivation. This gas will be called the ω-gas. We will denote the total thermodynamic entropy of a system by H, with a subscript which may indicate whether it is the total entropy of a gas, which potentially depends both on the states of the GPT systems in the boxes and on the classical degrees of freedom (positions, momenta) of the boxes, or just the entropy of the GPT or of the classical degrees of freedom.

At first we need to investigate how the entropy of the gas and of the ensemble are related to each other because later on, we will only consider the gas. So we consider also a second GPT ensemble $[S_1', \ldots, S_N']$ (described by $\omega' \in \Omega_A$) implanted into a gas the same way. At temperature $T = 0$, the movement of the boxes freezes out and we are left with the GPT ensembles. In this case, the thermodynamic entropies of the gases and the GPT ensembles must satisfy: $H_{\omega\text{-gas}} - H_{\omega'\text{-gas}} = H_{\omega\text{-ensemble}} - H_{\omega'\text{-ensemble}}$. Remember that the heat capacity is $C = \delta Q/\mathrm{d}T$, and as the gases only differ in their internal systems, which are isolated, C is the same for both gases. With $\mathrm{d}H = \delta Q/T$ we thus find that $H_{\omega\text{-gas}} - H_{\omega'\text{-gas}}$ is constant in T, i.e. $H_{\omega\text{-gas}} - H_{\omega'\text{-gas}} = H_{\omega\text{-ensemble}} - H_{\omega'\text{-ensemble}}$ for all temperatures.

The central tool for the thought experiment is a semi-permeable membrane. Whenever a box reaches the membrane, the membrane opens that box and measures the internal system. Depending on the result, a window is opened to let the box pass, or the window remains closed. It is crucial to note that this membrane will not cause problems in the style of Maxwell's demon, as was already discussed by von Neumann himself, because the membrane does not distinguish between its two sides.

Now we begin with the experiment itself; see figure 2. We consider a state $\omega = \sum_{j=1}^{n} p_j \omega_j$ where ω_j are perfectly distinguishable pure states, and $p_j = N_j/N$, where N_j boxes contain a system in the state ω_j. We assume that the ω-gas is confined in a container of volume V. Let there be a second container which is identical to the first one, but empty. The containers are merged together, the wall of the non-empty container separating the containers replaced by a semi-permeable membrane which lets only ω_1 pass. At the opposite wall of the non-empty container we insert a semi-permeable membrane which only blocks ω_1. The solid wall in the middle and the outer semi-permeable membrane are moved at constant distance until the solid wall hits the other end.

Once this is accomplished, i.e. in stage (4) in figure 2, one container has all ω_1-boxes and the other one contains all the rest. Note that this procedure is possible without performing any work as can be seen via Dalton's Law [90]: the work needed to push the semi-permeable membrane against the ω_1-gas can be recollected at the other side from the moving solid wall, which is pushed by the ω_1-gas into empty space. Thus we have separated the ω_1-boxes from the rest. We repeat a similar procedure until all the ω_j-gases are separated into separate containers of volume V.

Next we compress the containers isothermally to the volumes $p_j V$, respectively. Denoting the pressure by P, and using the ideal gas law, we obtain the required work

$$\int_V^{p_j V} P \, \mathrm{d}V = \int_V^{p_j V} N_j k_B T/V \, \mathrm{d}V = p_j N k_B T \log p_j,$$

where log denotes the natural logarithm. As the temperature and thus the internal energy remain constant, we extract heat $N k_B T \sum_j p_j \log p_j$.

At this point, we have achieved that every container contains a pure state ω_j. We now transform every ω_j to another pure state ω' which we choose to be the same for all containers. This is achieved by opening the boxes and applying a reversible transformation T_j in every container j which satisfies $T_j \omega_j = \omega'$. These transformations exist due to postulate 1. Since the same transformation T_j is applied to all small boxes in any given container j (without conditioning on the content of the small box), this operation is thermodynamically reversible.

Now we merge the containers, ending with a pure ω'-gas in the same condition as the initial ω-gas. This merging is reversible, because the density is not changed and because all states are the same, so one can just put in the walls again. The only step that caused an entropy difference was the isothermal compression. Thus, the difference of the entropies between the ω-gas and the ω'-gas (which are equal to the entropies of the respective GPT ensembles) is $N k_B \sum_j p_j \log p_j$. Therefore $H_{\omega\text{-ensemble}} = H_{\omega'\text{-ensemble}} - N k_B \sum_j p_j \log p_j$. If we assume that pure states have entropy zero, we thus end up with

$$H_{\omega\text{-ensemble}} = -N k_B \sum_j p_j \log p_j \tag{1}$$

[11] For a more detailed discussion of the physical properties of these small boxes, we refer the reader to von Neumann's original work [41].

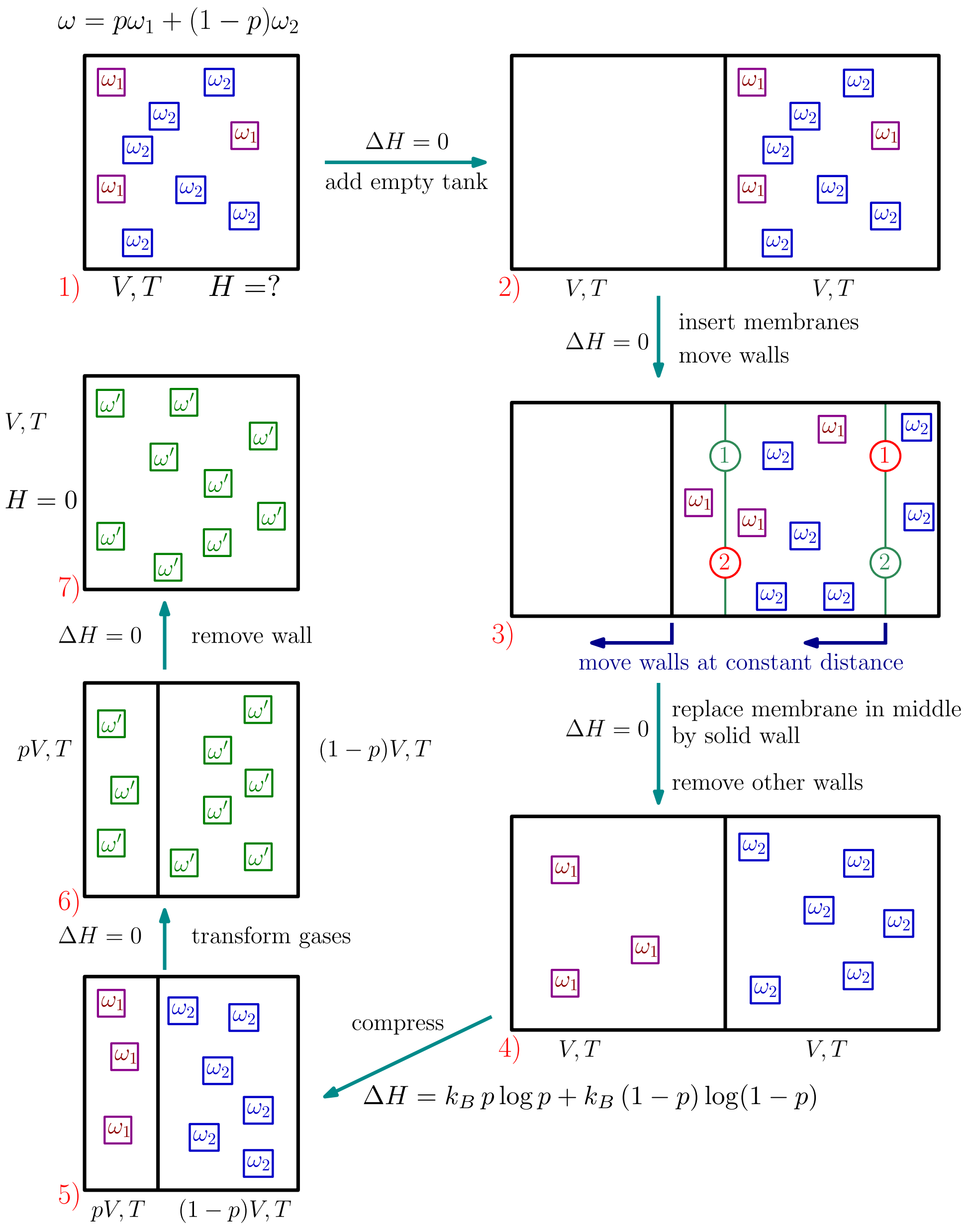

Figure 2. This figure shows von Neumann's thought experiment, as described in the main text. Stages (1)–(5) also feature in Petz' version.

and with the following entropy per system of the ensemble:

$$H(\omega) := \frac{1}{N} H_{\omega\text{-ensemble}} = -k_B \sum_j p_j \log p_j. \tag{2}$$

In summary, we have made the following assumptions to arrive at this notion of thermodynamic entropy:

Assumptions 1.

(a) Every (mixed) state can be prepared as an ensemble/statistical mixture of perfectly distinguishable pure states.

(b) A measurement that perfectly distinguishes those pure states can be implemented as a semi-permeable membrane, which in particular does not disturb the pure states that it distinguishes.

(c) All pure states can be reversibly transformed into each other.

(d) Thermodynamical entropy H is continuous in the state. (Since ensembles must have rational coefficients $p_j = N_j/N$, we need this to approximate arbitrary states in the thought experiment.)

(e) All pure states have entropy zero.

A generalised version of the thought experiment presented by Petz [42] is applicable to more general decompositions: suppose that $\omega_1, \ldots, \omega_n \in \Omega_A$ are perfectly distinguishable, but not necessarily pure. Let $p_1, \ldots, p_n$ be a probability distribution. Then Petz' thought experiment implies that

$$H\left(\sum_j p_j \omega_j\right) = \sum_j p_j H(\omega_j) - k_B \sum_j p_j \log p_j. \tag{3}$$

The main idea is that steps (1)–(5) of von Neumann's thought experiment can be run even if the perfectly distinguishable states $\omega_1, \ldots, \omega_n$ are mixed and not pure (as long as the membrane will still keep them undisturbed). Then the entropy of the state in (5) can be computed by making an additional *extensivity assumption*: denote the GPT entropy of an ω-ensemble of N particles in a volume V by $H_{\omega\text{-ensemble}}(N, V)$, then this assumption is that

$$H_{\omega\text{-ensemble}}(\lambda N, \lambda V) = \lambda\, H_{\omega\text{-ensemble}}(N, V)$$

for $\lambda \geqslant 0$. Assuming in addition that the entropy of the n containers adds up, the total entropy of the configuration in step (5) is $N \sum_j p_j H(\omega_j)$, from which Petz obtains (3). While this approach needs this additional extensivity assumption, it does not need to postulate that all pure states can be reversibly transformed into each other (in contrast to von Neumann's version). Under the assumption that all pure states have entropy zero, it reproduces equation (2) as a special case.

We conclude this section with a few comments on the idealisations used in the thought experiments above. The use of gases in which the exact numbers of particles with each internal state is known parallels von Neumann's argument in [41]. We rarely if ever have such precise knowledge of particle numbers in real physical gases, so our argument involves a strong idealisation, but one that is common in thermodynamics and that has also been made by von Neumann[12].

Although fluctuations in work are significant for small particle numbers, in the thermodynamic limit of large numbers of particles there is concentration about the expected value given, in von Neumann's protocol, by the von Neumann entropy, and therefore our arguments (and von Neumann's) have the most physical relevance in this large-N situation. This is of course true for classical thermodynamics as well—indeed, the use made of the ideal gas law and Dalton's law in von Neumann's argument are additional places where large N is needed if one wants fluctuations to be negligible. We expect finer-grained considerations to be required for a thorough study of fluctuations in finite systems, which is one reason for interest in the additional entropic measures studied in section 5.6, but von Neumann's argument does not concern these finer-grained aspects of the thermodynamics of finite systems.

4. Why the 'gbit' is ruled out

In section 2, we have introduced the 'gbit', a system for which the state space Ω is a square. Gbits are particularly interesting because they correspond to 'one half' of a Popescu–Rohrlich box [83] which exhibits correlations that are stronger than those allowed by quantum theory [70]. One might wonder whether the thought experiments of section 3 allow us to define a notion of thermodynamic entropy for the gbit. We will now show that this is not the case, which can be seen as a thermodynamical argument for why we do not see superstrong correlations of the Popescu–Rohrlich type in our universe.

Since not all states of a gbit can be written as a mixture of perfectly distinguishable *pure* states, von Neumann's original thought experiment cannot be of direct use here. However, we may resort to Petz' version: every mixed state ω of a gbit can be written as a mixture of perfectly distinguishable *mixed* states, as illustrated in figure 3. Furthermore, the other crucial assumption on the state space is satisfied, too: for every pair of perfectly distinguishable mixed states, there is an instrument (a 'membrane') that distinguishes those states without disturbing them. We even have that all pure states can be reversibly transformed into each other (namely by a rotation of the square).

Thus, we can analyse the behaviour of a gbit state space in Petz' version of the thought experiment. Any continuous notion of thermodynamic entropy H consistent with this thought experiment would thus have to

[12] Here, von Neumann's thought experiment is formulated in terms of a frequentist view on probabilities, which is standard in most treatments on thermodynamics. A treatment involving a finite ensemble where the frequencies (and perhaps the total particle number) are stochastic might seem more suitable from a Bayesian point of view; it would likely raise issues about whether the amount of work extracted from a finite system is subject to fluctuations. For systems that are finite or out of equilibrium, measures such as Shannon's are known not to be the whole story (see [30] and references therein). But even for finite systems with a more realistic treatment of uncertainty about particle numbers, the von Neumann entropy still gives the *expected* work in the protocol he considers. We defer these issues to future work, although we note that [30] suggests the operational entropies discussed in section 5.6 are among the relevant tools for tackling them.

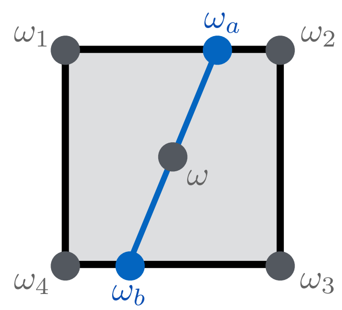

Figure 3. In an attempt to define a notion of thermodynamic entropy for the gbit, we can decompose any state into perfectly distinguishable states. This is done in two steps, as explained in the main text.

satisfy (3). However, we will now show that the gbit does not admit any notion of entropy that satisfies (3). Consider different decompositions of the state $\omega = \frac{1}{2}\omega_a + \frac{1}{2}\omega_b$ in the centre of the square, where $\omega_a = p\omega_1 + (1 - p)\omega_2$ as well as $\omega_b = p\omega_3 + (1 - p)\omega_4$. It is geometrically clear that every choice of $0 < p < 1$ corresponds to a valid decomposition. We find (applying equation (3) to ω for the first equality, and to ω_a and ω_b for the second):

$$H(\omega) = \frac{1}{2}H(\omega_a) + \frac{1}{2}H(\omega_b) - 2k_B\frac{1}{2}\log\frac{1}{2} = \frac{1}{2}p\,H(\omega_1) + \frac{1}{2}(1-p)H(\omega_2) + \frac{1}{2}p\,H(\omega_3) + \frac{1}{2}(1-p)H(\omega_4) - k_B p\log p - k_B(1-p)\log(1-p) + k_B\log 2.$$

This expression can never be constant in p, no matter what value of entropy of the four pure states $H(\omega_i)$ we assume. Thus, the entropy $H(\omega)$ of the centre state ω is not well-defined, since it depends on the choice of decomposition.

In other words, the structure of the gbit state space enforces that any meaningful notion of thermodynamic entropy H will not only be a function of the *state*, but a function of the *ensemble that represents the state*. If a state ω is represented by different ensembles, then this will in general give different values of entropy.

So what goes wrong for the gbit? Clearly, all we can say with certainty is that the combination of assumptions made in von Neumann's thought experiment turns out not to yield a unique notion of entropy, while a deeper physical interpretation seems only possible under further assumptions on the interplay between the gbit and the thermodynamic operations. However, a comparison with quantum theory motivates at least one further speculative attempt at interpretation. In the example above, we have decomposed a state ω into two perfectly distinguishable states ω_a and ω_b, which can themselves be decomposed into pairs of perfectly distinguishable states ω_1 and ω_2, or ω_3 and ω_4 respectively. In quantum theory, this would only be possible if ω_a and ω_b are orthogonal, which would then imply that all four states $\omega_1, \ldots, \omega_4$ are pairwise orthogonal. This would enforce that there exists a unique projective measurement (a 'membrane') that distinguishes all these four states jointly. This membrane could feature in von Neumann's thought experiment (or other similar thermodynamical settings), yielding a unique notion of thermodynamic entropy.

On the other hand, in the gbit, the four pure states $\omega_1, \ldots, \omega_4$ are *not* jointly perfectly distinguishable. Hence there is no canonical choice of 'membrane' that could be used in the thought experiment to define a unique natural notion of entropy for the gbit states. Entropy will be 'contextual', depending on the choice of membrane resp. ensemble decomposition that is used in any given specific thermodynamical setting. Therefore, the implication '*pairwise distinguishability* $\Rightarrow$ *joint distinguishability*', which is true for quantum theory, has thermodynamic relevance. This implication, if suitably interpreted, leads to the 'exclusivity principle' [7, 8, 91], namely that the sum of the probabilities of pairwise exclusive propositions cannot exceed 1 (in this case these propositions correspond to the outcomes of the jointly distinguishing measurement). This suggests that the exclusivity principle, which has so far been considered only in the realm of contextuality, may be thermodynamically relevant. This observation is also closely related to the notion of 'dimension mismatch' described in [82], and to orthomodularity in quantum logic (see for example [23]).

5. A class of theories with consistent thermodynamic behaviour

5.1. The two postulates

In this section we introduce the two postulates that express key operational concepts from thermodynamics. The first postulate is motivated by the universality of thermodynamics and the distinction between microscopic and macroscopic behaviour. At first we consider the universality of thermodynamics, in the sense that thermodynamics is a very general theory whose basic principles can be applied to many possible implementations, as already noticed by N Carnot [44]:

'In order to consider in the most general way the principle of the production of motion by heat, it must be considered independently of any mechanism or any particular agent. It is necessary to establish principles applicable not only to steam engines but to all imaginable heat-engines, whatever the working substance and whatever the method by which it is operated.'

Recalling von Neumann's thought experiment in the case of quantum theory, we can think of thermodynamical protocols (which will ultimately also include heat engines) as acting on a given ensemble, defined as a probabilistic mixture of pure states chosen from a fixed basis. If we interpret ensembles with different choices of basis as different 'working substances', then Carnot's principle should apply: protocols that can be implemented on one ensemble (say, ensemble 1) can also be implemented on the other (say, ensemble 2)[13]. In quantum theory, this universality is ensured by the existence of unitary transformations: all orthonormal bases can be translated into each other by a unitary and therefore reversible map. In this sense, the state of ensemble 1 can in principle be transferred to ensemble 2, then the thermodynamic protocol of ensemble 2 can be performed (if we have also transformed the projectors describing the membranes accordingly), and then one can transform back. Even if this cannot always be achieved in practice, the corresponding unitary symmetry of the quantum state space (considered as passive transformations between different descriptions) enforces the aforementioned universality[14].

This universality of implementation, as well as independence of the choice of labels and descriptions, should continue to hold in all generalised theories that we consider. An orthonormal basis from quantum theory is nothing else than a set of perfectly distinguishable pure states, i.e. an n-frame. Therefore, in our generalised theories, we expect that this universality of implementation is achieved by the existence of reversible transformations that, in analogy to unitary maps, transform any given n-frame into any other:

Postulate 1. For each $n \in \mathbb{N}$, all sets of n perfectly distinguishable pure states are equivalent. That is, if $\{\omega_1, \ldots, \omega_n\}$ and $\{\varphi_1, \ldots, \varphi_n\}$ are two such sets, then there exists a reversible transformation T with $T\omega_j = \varphi_j$ for all j.

Furthermore, postulate 1 expresses a physical property that is crucial for thermodynamics: that of *microscopic reversibility*. Many characteristic properties of thermodynamics arise from limited experimental access to the microscopic degrees of freedom, which by themselves undergo reversible time evolution. This reversibility, for example, forbids evolving two microstates into one, which is at the heart of the non-decrease of entropy. If the experimenter had full access to the microscopic degrees of freedom, then he or she could convert any state of maximal knowledge to any other one as long as he or she preserved distinguishability. Postulate 1 formalises this microscopic basis of thermodynamics by demanding the existence of 'enough' distinguishability-preserving, microscopic transformations T, which can be understood as reversible time evolutions.

Postulate 1 has substantial information-theoretical justifications and consequences. The basic concepts of both thermodynamics and information processing are independent of the choice of implementation. For information processing this is formalised by the Turing machine which admits a multitude of physical realisations. Perfectly distinguishable pure states can be taken as bits, and postulate 1 expresses that all bits (or their higher-dimensional analogues) are equivalent. It is for this reason that postulate 1 was called *generalised bit symmetry* in [34], and its restriction to pairs of distinguishable states was called *bit symmetry* in [64]. Starting with Landauer's principle, 'thermodynamics of computation' [92] has become a fruitful paradigm that relates the two apparently disjoint fields. The two complementary interpretations of postulate 1 are one instance of this.

Now we turn to our second postulate. We are looking for theories very similar to the thermodynamics we are used to; thus it is essential that we can adopt basic notions of standard thermodynamics unchanged or with only very small alterations. Two such notions of great importance are (Shannon) entropy $S = -k_B \sum_j p_j \log p_j$ and majorization theory. In classical and quantum thermodynamics, these notions operate on the coefficients in a decomposition of a state into perfectly distinguishable pure states (in quantum theory, the eigenvalues). In order to not change thermodynamic theory too much, we would also like this to be possible in our more general state spaces. Thus, we demand that every state has a convex decomposition into perfectly distinguishable pure states.

[13] Here we only consider ensembles of identical Hilbert space dimensions. If the dimensions are different (say, 2 versus 3), then one can implement different sets of protocols on the ensembles (say, ones involving semipermeable membranes that distinguish 3 alternatives in the latter, but not the former case). One could then still discuss a notion of universality in Carnot's spirit, by referring to the equivalence of, say, a state space with $N = 3$ alternatives to a subspace of a state space with $N = 2 \times 2$ alternatives, but we will not discuss this further here.

[14] In classical thermodynamics, the analogue of a choice of basis is the labelling of the distinguishable configurations. Clearly, the availability of thermodynamic protocols does not change under relabelling.

Note that this was indeed one of our assumptions in von Neumann's thought experiment in section 3. There, it allowed us to realise any state ω as a 'quasiclassical ensemble', i.e. as an ensemble of states that behave like classical labels. This gives us a further justification of our second postulate: thermodynamic (thought) experiments require that states have an ensemble interpretation. An unambiguous notion of 'counting of microstates' demands that the ensembles consist of perfectly distinguishable, pure states. Without this, obtaining a phenomenological thermodynamics for which the theory is the underlying microscopic theory seems problematic. Thus, our second postulate is

Postulate 2. Every state $\omega \in \Omega_A$ has a convex decomposition $\omega = \sum_j p_j \omega_j$ into perfectly distinguishable pure states ω_j.

It is tempting to interpret the two postulates as reflecting the *microscopic* and the *macroscopic* aspects of thermodynamics, respectively: while postulate 1 describes microscopic reversibility of the pure states that may describe single particles in thermodynamics, postulate 2 ensures that mixed states can be interpreted macroscopically as descriptions of quasiclassical ensembles, composed of a large number of particles that are separately in unknown but distinguishable microstates.

We will not introduce any further postulates. In particular, we will not make any assumptions on the *composition* of systems. All our results are therefore independent from notions like *tomographic locality* [51] (which is arguably dispensable in many important situations [81]) or *purification* [56] (which is a rather strong assumption); we do not assume either of the two.

5.2. Some consequences of postulates 1 and 2

Postulates 1 and 2 have been analysed in [43], but in a different context: instead of investigating thermodynamics, the goal in [43] was to obtain a reconstruction of quantum theory, by supplementing postulates 1 and 2 with further postulates. Some of the insights from [43] will be important here, and are therefore briefly discussed below. Starting with section 5.4, we will also obtain new results that are interesting in a thermodynamic context.

In contrast to Hilbert space, there is no apriori notion of inner product for GPTs. However, as shown in [64], we get a natural inner product $\langle \cdot, \cdot \rangle$ as a consequence of postulates 1 and 2: it satisfies $\langle T\omega, T\varphi \rangle = \langle \omega, \varphi \rangle$ for all reversible transformations T, and $0 \leqslant \langle \omega, \varphi \rangle \leqslant 1$ for all states $\omega, \varphi \in \Omega_A$. Furthermore, $\langle \omega, \omega \rangle = 1$ for all pure $\omega \in \Omega_A$ and $\langle \varphi, \varphi \rangle < 1$ for all mixed $\varphi \in \Omega_A$, and $\langle \omega, \varphi \rangle = 0$ if $\omega, \varphi \in \Omega_A$ are perfectly distinguishable. Thus, all perfectly distinguishable states are orthogonal, as in quantum theory.

Moreover, the cone of unnormalized states becomes *self-dual* with this choice of inner product. In particular, every effect e can be taken as a vector in A_+, such that $e(\omega) = \langle e, \omega \rangle$. In standard quantum theory, this is the Hilbert Schmidt inner product on the real vector space of Hermitian matrices: $\langle X, Y \rangle = \mathrm{tr}(XY)$ for $X = X^\dagger, Y = Y^\dagger$.

Quantum theory has more structure: the convex set of density matrices Ω_A has faces[15], and these faces are in one-to-one correspondence to subspaces of Hilbert space (namely, a face F contains all density matrices that have support on the corresponding Hilbert subspace). To every face F, we can associate a number $|F|$ which is the dimension of the corresponding Hilbert subspace, and $F \subsetneq G$ implies $|F| < |G|$. Every face F is generated by $|F|$ pure and perfectly distinguishable states in F (an $|F|$-frame in F), and every (smaller) frame that is a subset of F can be completed, or extended, to a frame which has $|F|$ elements and thus generates F.

In all theories that satisfy postulates 1 and 2, all these properties hold in complete analogy [43]. However, since faces do not any more correspond to Hilbert spaces, the numbers $|F|$ do not have an interpretation as the dimension of a subspace. Instead, we call $|F|$ the *rank* of F. If von Neumann's thought experiment is supposed to make sense for these theories, we need a way to formalise the working of a semipermeable membrane, which in quantum theory is done via projective measurements.

Since we are dealing with unnormalized states, the corresponding analogue in GPTs will be formulated in terms of the set of unnormalized states A_+. As one can see in the case of the gbit, it is not automatic that we have any notion of 'projective measurements' for any given state space. However, postulates 1 and 2 turn out to ensure that projective measurements exist. For any face F of A_+ (the non-negative multiples of the corresponding face of Ω_A), consider the orthogonal projector P_F onto the span of F. One can show that P_F is *positive*, i.e. maps (unnormalized) states to (unnormalized) states [43]. Moreover, P_F does not disturb the states in the face F.

Thus, to a given set of mutually orthogonal faces $F_1, \ldots, F_m$ such that $|F_1| + \ldots + |F_m| = N_A$, we can associate an instrument with transformations $T_i := P_{F_i}$, which describes a projective measurement, as in a

[15] A *face* of a convex set C is a convex subset $F \subseteq C$ with the property that $\lambda x + (1-\lambda)y \in F$ with $0 < \lambda < 1$ and $x, y \in C$ implies $x, y \in F$ [89]. We say that F is *generated by* $\omega_1, \ldots, \omega_n$ if F is the smallest face that contains $\omega_1, \ldots, \omega_n$.

semipermeable membrane. Transformation T_i leaves the states in face F_i unperturbed, but fully blocks out states in the other faces, i.e. $T_i\omega = 0$ for $\omega \in F_j$, $i \neq j$. In standard quantum theory, these transformations are $P_{F_i}\rho = \pi_i\rho\pi_i$, where π_i is the orthogonal Hilbert space projector onto the ith Hilbert subspace. The rank condition becomes $\mathrm{tr}(\pi_1) + \ldots + \mathrm{tr}(\pi_m) = N_A$ (the total Hilbert space dimension), and mutual orthogonality is $\pi_i\pi_j = \delta_{ij}\pi_i$. We will show in section 5.4 that the mutually orthogonal faces replace the eigenspaces from quantum theory and that the projective measurement described here can be interpreted as measuring an observable.

The Hilbert space projector π_i therefore also has an interpretation as an *effect* in standard quantum theory: it yields the probability of outcome i in the projective measurement on a state ρ, namely $\mathrm{tr}(\pi_i\rho)$. The analogous effect in a GPT that satisfies postulates 1 and 2, corresponding to a face F, is

$$u_F := P_F u_A$$

(identifying the effect u_A with a vector via the inner product). The effect u_F is sometimes called the 'projective unit' of F. In quantum theory, we can write $\pi_i = \sum_j |\psi_j\rangle\langle\psi_j|$, where the $|\psi_j\rangle$ are an orthonormal basis of the corresponding Hilbert subspace. The same turns out to be true in our GPTs: we have

$$u_F = \sum_{j=1}^{|F|}\omega_i, \tag{4}$$

where $\omega_1, \ldots, \omega_{|F|}$ is any frame that generates F. Therefore, the probability to obtain outcome i in the projective measurement above on state ω is $\langle u_{F_i}, \omega\rangle = \langle u_A, P_{F_i}\omega\rangle$.

5.3. State spaces satisfying postulates 1 and 2

It is easy to see that both quantum and classical state spaces satisfy postulates 1 and 2. By a 'classical state space', we mean a state space that consists of discrete probability distributions. Concretely, for any number $N \in \mathbb{N}$ of mutually exclusive alternatives, consider the state space

$$\Omega := \{(p_1, \ldots, p_N) \mid p_i \geqslant 0, \ \sum_i p_i = 1\}.$$

Any pure state is given by a deterministic probability vector, i.e. $\omega_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ (where 1 is on the ith place). If we have two equally sized sets of such vectors (as in postulate 1), then there is always a permutation that maps one set to the other. In fact, the reversible transformations correspond to the permutations of the entries. Postulate 2 is then simply the statement that

$$(p_1, \ldots, p_N) = p_1\omega_1 + p_2\omega_2 + \ldots + p_N\omega_N.$$

Which state spaces are there, in addition to standard complex quantum theory and classical probability theory, that satisfy postulates 1 and 2? We think that this question is very difficult to answer. Thus, we formulate the following

Open problem 1. Classify all state spaces that satisfy postulates 1 and 2.

From the results in [43], we know which state spaces satisfy postulates 1 and 2 *and* one additional property: the absence of third-order interference. The notion of higher-order interference has been introduced by Sorkin [76], and has since been the subject of intense theoretical [93, 95, 96] and experimental [97–102] interest.

For the main idea, think of three mutually exclusive alternatives in quantum theory (such as three slits in a triple-slit experiment), described by orthogonal projectors π_1, π_2, π_3. The event that alternative 1 *or* alternative 2 takes place is described by the projector $\pi_{12} = \pi_1 + \pi_2$; similarly, we have π_{13}, π_{23} and π_{123}. Their actions on density matrices are described by superoperators

$$\rho \mapsto P_{12}(\rho) := \pi_{12}\rho\pi_{12}$$

(and similarly for the other projectors). As a consequence, we obtain that $P_{12} \neq P_1 + P_2$, which expresses the phenomenon of *interference*. However, it is easy to check that

$$P_{123} = P_{12} + P_{13} + P_{23} - P_1 - P_2 - P_3, \tag{5}$$

which means that interference over three alternatives can be reduced to contributions from interferences of pairs of alternatives. Similar identities hold for an arbitrary number $n \geqslant 4$ of alternatives: *quantum theory admits only pairwise interference*, and no 'third-order interference' which would be characterised by a violation of this equality.

In the context of postulates 1 and 2, we have an analogous notion of orthogonal projectors, and thus we can consider (5) and its generalisation to $n \geqslant 4$ alternatives on a state space with $N \geqslant n$ perfectly distinguishable states. Postulating this 'absence of third-order interference' in addition to postulates 1 and 2 gives us the following:

Theorem 2 (Lemma 33 in [43]). *The possible state spaces which satisfy postulates 1 and 2 and which do not admit third-order interference, in addition to classical state spaces, are the following. First, for $N \geqslant 4$ perfectly distinguishable states, there are only three possibilities:*

- *Standard complex quantum theory.*
- *Quantum theory over the real numbers. That is, only real entries are allowed in the $N \times N$ density matrices.*
- *Quantum theory over the quaternions. The state spaces are the self-adjoint $N \times N$ quaternionic matrices of unit trace.*

For $N = 3$ perfectly distinguishable states, all of the above and one exceptional solution are possible, namely quantum theory over the octonions (but only for the case of 3×3 unit trace density matrices).

For $N = 2$ (the 'bit' case), we have the d-dimensional Bloch ball state spaces $\Omega_d := \{(1, r)^T | r \in \mathbb{R}^d, \|r\| \leqslant 1\}$ with $d \geqslant 2$. They are analogous to the standard Bloch ball Ω_3 of quantum theory, with very similar descriptions of effects etc. Their group of reversible transformations may either be $\mathrm{SO}(d)$ (which corresponds to $\mathrm{PU}(2)$ for $d = 3$), or some subgroup of $\mathrm{O}(d)$ which is transitive on the sphere (such as $\mathrm{SU}(2)$ for $d = 4$).

Mathematically, these examples correspond to the state spaces of the finite-dimensional irreducible formally real Jordan algebras [24, 43]. We do not know whether there are theories that satisfy postulates 1 and 2 but admit higher-order interference and therefore do not appear on this list. In theorem 12, we will show that the question whether a theory has third-order interference is related to the properties of its Rényi entropies.

5.4. Observables and diagonalization

A central part of physics are observables and how they can be measured. In standard quantum theory, we can introduce observables in two different ways, which both equivalently lead to the prescription that *observables are described by Hermitian operators/matrices*.

First, in finite dimensions, we can characterise observables as *those objects that linearly assign real expectation values to states*. In the case of quantum theory it follows that observables are represented by matrices X, and Hermiticity $X = X^\dagger$ implies that expectation values $\mathrm{tr}(\rho X)$ are always real. Linearity is enforced by the statistical interpretation of states, for the same reason that effects in GPTs are linear.

Second, we can introduce observables by saying that there is a projective measurement $\pi_1, \ldots, \pi_n$ that measures this observable, and which has outcomes $x_1, \ldots, x_n \in \mathbb{R}$. This leads to the Hermitian operator $X = \sum_{i=1}^n x_i \pi_i$. Since every Hermitian operator can be diagonalized, these two definitions are equivalent.

Our two postulates provide the structure to introduce observables in a completely analogous way. First, using the inner product, we can define observables as linear maps of the form

$$\omega \mapsto \langle x, \omega \rangle$$

and thus identify them with elements $x \in A$ of the vector space that carries the states (as in quantum theory, where this vector space is the space of Hermitian matrices). As noticed in [62], every such vector has a representation of the form

$$x = \sum_i x_i u_i, \tag{6}$$

where the u_i are projective units corresponding to mutually orthogonal faces F_i, $x_i \in \mathbb{R}$, and $x_i \neq x_j$ for $i \neq j$. The analogy with quantum theory goes even further: due to (4), we have $x = \sum_i x_i \sum_j \omega_i^{(j)}$, whenever $\omega_i^{(1)}, \ldots, \omega_i^{(|F_i|)}$ is a frame on F_i. This corresponds to the identity $X = \sum_i x_i \sum_j |\psi_i^{(j)}\rangle\langle\psi_i^{(j)}|$ in standard quantum theory. In analogy to quantum theory we will call the F_i eigenfaces and the x_i eigenvalues. To further justify this terminology, note that the x_i are eigenvalues of the map $\sum_i x_i P_i$, where P_i are the orthogonal projectors onto the spans of the faces F_i.

Theorem 3. *If postulates 1 and 2 hold, then every element $x \in A$ has a representation of the form $x = \sum_{j=1}^n x_j u_j$ where $x_j \in \mathbb{R}$ are pairwise different and the u_j are the projective units of pairwise orthogonal faces F_j such that $\sum_j u_j = u_A$. This decomposition $x = \sum_{j=1}^n x_j u_j$ is unique up to relabelling. In analogy to quantum theory, we will call the x_j eigenvalues and the F_j eigenfaces.*

Furthermore, for every real function f with suitable domain of definition, we can define

$$f(x) := \sum_{j=1}^n f(x_j) u_j \tag{7}$$

as in spectral calculus.

If P_j is the orthogonal projector onto the span of F_j, then $(P_1, \ldots, P_n)$ is a well-defined instrument with induced measurement $(u_1, \ldots, u_n)$ which leaves the elements of $\mathrm{span}(F_j)$ *invariant:*

$$P_j(\omega) = \delta_{jk} \cdot \omega \quad \text{for all } \omega \in F_k.$$

In analogy to quantum theory, we will call this instrument the projective measurement of the observable x.

We will give a proof in the appendix[16]. Equation (7) allows us to define a notion of entropy, in full analogy to quantum mechanics.

Definition 4 (Spectral entropy). If A is a state space that satisfies postulates 1 and 2, we define the spectral entropy for any state $\omega \in \Omega_A$ as

$$S(\omega) := -\sum_i p_i \log p_i,$$

where $\omega = \sum_i p_i \omega_i$ is any convex decomposition of ω into pure and perfectly distinguishable states ω_i, and $0 \log 0 := 0$.

Theorem 3 tells us that this definition is independent of the choice of decomposition: it is easy to check that

$$S(\omega) = -\langle \omega, \log \omega \rangle,$$

where $\log \omega$ is understood in the sense of spectral calculus as in (7). The right-hand side is manifestly independent of the decomposition. It can also be written $S(\omega) = u_A(\eta(\omega))$, where $\eta(x) = -x \log x$ for $x > 0$ and $\eta(0) = 0$. In particular,

$$\omega \text{ is a pure state} \Leftrightarrow S(\omega) = 0. \tag{8}$$

To see this, note that any pure state $\omega_1 = \omega$ can be extended to a set of perfectly distinguishable pure states $\omega_1, \omega_2, \ldots, \omega_{N_A}$ such that $\omega = 1 \cdot \omega_1 + 0 \cdot \omega_2 + \ldots + 0 \cdot \omega_{N_A}$. Conversely, if $S(\omega) = 0$, then any decomposition of ω must have coefficients $(1, 0, \ldots, 0)$.

5.5. Thermodynamics in the context of postulates 1 and 2

If a state space satisfies postulates 1 and 2, then it also satisfies all the assumptions that we have made in von Neumann's thought experiment. It is easy to check all items in assumptions 1: (a) is simply postulate 2, and (c) is a consequence of postulate 1. As we have seen in the previous section, our two postulates imply that we have orthogonal projectors sharing important properties with those of standard quantum theory. If we make the physical assumption that we can actually implement them by means of semipermeable membranes (as in quantum theory), we obtain (b). Item (e) is the same as (8). Note that assumption (d) is not a mathematical assumption about the state space, but a physical assumption about thermodynamic entropy. This shows part of the following (the full proof will be given in the appendix):

Observation 5. von Neumann's thought experiment, as explained in section 3, can be run for every state space that satisfies postulates 1 and 2. The notion of thermodynamic entropy H that one obtains from that thought experiment turns out to equal spectral entropy S as given in definition 4,

$$H(\omega) = S(\omega) \qquad \text{for all states } \omega.$$

This is consistent with assumptions 1. Furthermore, it is also consistent with Petz' version of the thought experiment, because spectral entropy satisfies

$$S(\omega) = \sum_j p_j S(\omega_j) - \sum_j p_j \log p_j \tag{9}$$

for every convex decomposition $\omega = \sum_j p_j \omega_j$ of ω into perfectly distinguishable, not necessarily pure states ω_j.

Thus, spectral entropy S gives meaningful and consistent physical predictions in situations like von Neumann's and Petz' thought experiments. However, we clearly do not know whether S is a consistent notion of physical entropy in *all* thermodynamical situations.

It turns out that there are further properties of S that encourage its physical interpretation as a thermodynamical entropy. In particular, we will now show that the *second law* holds in two important situations. We start by considering projective measurements $P_1, \ldots, P_n$. Projective measurements can model semipermeable membranes as in von Neumann's thought experiment, or they describe the measurement of an observable as

[16] This can also also obtained by combining the fact that postulates 1 and 2 imply the state space is projective (first part of theorem 17 in [43]) and self-dual (proposition 3 in [43]) with results such as theorem 8.64 in [24].

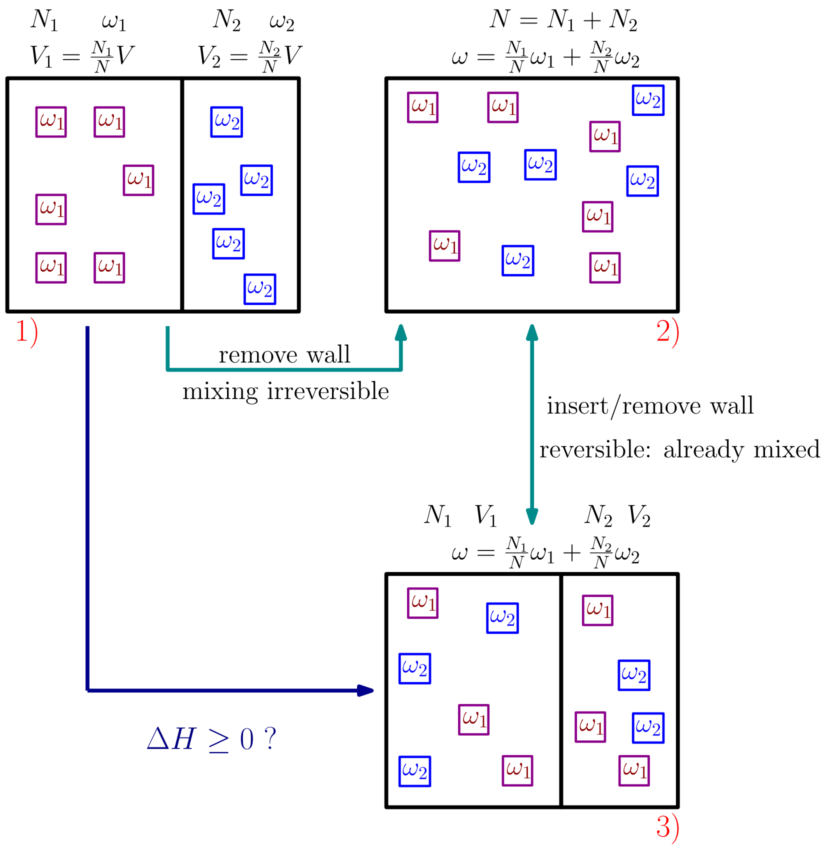

Figure 4. A process mixing gases by removing a separating wall. Theorem 7 ensures that this process does not decrease entropy, i.e. $\Delta H \geqslant 0$, if thermodynamic entropy H is identified with spectral entropy S as suggested by von Neumann's thought experiment.

explained in section 5.4. Consider the action of this measurement on a given state ω. With probabilities $(u_A \circ P_j)(\omega)$, this measurement yields the outcome j with post-measurement state $\omega_j := P_j\omega/(u_A \circ P_j(\omega))$. Performing this measurement on every particle of an ensemble (without learning the outcomes) yields a new ensemble, described by the post-measurement state

$$\omega' = \sum_{j:\ u_A \circ P_j(\omega) \neq 0} (u_A \circ P_j)(\omega) \cdot \omega_j = \sum_j P_j \omega.$$

Projective measurements do not decrease the entropy of the ensemble:

Theorem 6. *Suppose postulates 1 and 2 are satisfied. Let $P_1, \ldots, P_n$ be orthogonal projectors which form a valid instrument. Then the induced measurement with post-measurement ensemble state $\omega' = \sum_j P_j w$ does not decrease entropy: $S(\omega') \geqslant S(\omega)$.*

The proof will be given in the appendix. As in standard quantum theory, projectors P_j form a valid instrument if and only if they are mutually orthogonal, i.e. $P_i P_j = \delta_{ij} P_i$, and complete: $\sum_i u_A \circ P_i = u_A$.

Another important manifestation of the second law is in mixing procedures as in figure 4. Consider tanks that are separated by walls. Similarly to von Neumann's thought experiment, let the jth tank contain an N_j-particle gas that represents an ω_j-ensemble. Furthermore, assume that all the gases are at the same pressure and density. Identifying thermodynamic entropy H with spectral entropy S (as suggested by observation 5), the entropy of the GPT-ensemble in tank j is $N_j S(\omega_j)$, where S is the entropy per system. Thus the total GPT-entropy is $\sum_j N_j S(\omega_j)$. We remove the walls and let the gases mix. Then we put the walls back in. Now all the tanks contain gases hosting $\sum_j \frac{N_j}{N}\omega_j$ ensembles at the same conditions as before, where $N = \sum_j N_j$. The total GPT-entropy in the end is given by $\sum_j N_j S\left(\sum_k \frac{N_k}{N}\omega_k\right) = NS\left(\sum_k \frac{N_k}{N}\omega_k\right)$. As the gases in the tanks have the same density, volume, temperature and pressure as before, the only difference in entropy is due to the GPT-ensembles. The second law requires that the entropy does not decrease in this process, i.e. that $\sum_j N_j S(\omega_j) \leqslant NS\left(\sum_j \frac{N_j}{N}\omega_j\right)$ and thus $\sum_j \frac{N_j}{N} S(\omega_j) \leqslant S\left(\sum_j \frac{N_j}{N}\omega_j\right)$. The following theorem shows that our two postulates guarantee that this is true:

Theorem 7. *Assume postulates 1 and 2. Then entropy is concave, i.e. for $\omega_1, \ldots, \omega_n \in \Omega_A$ and $p_1, \ldots, p_n$ a probability distribution, we have*

$$S\left(\sum_j p_j w_j\right) \geqslant \sum_j p_j S(w_j). \tag{10}$$

Thus, the second law automatically holds for mixing processes. One way to prove (10) is to see that S equals 'measurement entropy' as we will show in section 5.6, proven to be concave in [66, 67]. However, there is a simpler proof that uses a notion of *relative entropy*, which is an important notion in its own right.

Definition 8. For state spaces A that satisfy postulates 1 and 2, we define the *relative entropy* of two states $\omega, \varphi \in \Omega_A$ as

$$S(\omega\|\varphi) := -S(\omega) - \langle \omega, \log \varphi \rangle.$$

Here, for $\varphi = \sum_j q_j \varphi_j$ any decomposition into a maximal frame, $\log \varphi = \sum_j \log(q_j)\varphi_j$ according to theorem 3. (As in quantum theory, this can be infinite if there are $q_j = 0$ such that $\langle \omega, \varphi_j \rangle \neq 0$).

A notion of relative entropy in GPTs has also been defined in Scandolo's Master Thesis [48], but under different assumptions, as discussed in the introduction. Relative entropy continues to satisfy *Klein's inequality*, a fact that is useful in proving theorem 7. The proof is similar to that within standard quantum theory and deferred to the appendix.

Theorem 9 (Klein's inequality). *For all $\omega, \varphi \in \Omega_A$,*

$$S(\omega\|\varphi) \geqslant 0.$$

Klein's inequality can be used to give a simple proof of theorem 7:

$$0 \leqslant \sum_j p_j S\left(\omega_j \| \sum_k p_k \omega_k\right) = -\sum_j p_j S(\omega_j) - \left\langle \sum_j p_j \omega_j, \log\left(\sum_k p_k \omega_k\right)\right\rangle = -\sum_j p_j S(\omega_j) + S\left(\sum_k p_k \omega_k\right).$$

Given all the calculations in this subsection in terms of orthogonal projections, it may seem at first sight as if every statement or calculation in quantum theory can be analogously made in the more general state spaces that satisfy postulates 1 and 2. However, this may not quite be true, as the fact that the following is an open problem shows:

Open problem 2. For state spaces satisfying postulates 1 and 2, if ω is a pure state, and P an orthogonal projection, then is $P\omega$ also (up to normalisation) a pure state?

In classical and quantum state spaces, the answer is 'yes', but we do not know if a positive answer follows from postulates 1 and 2 alone. We will return to this problem in theorem 12.

Note that Chiribella and Scandolo have applied similar techniques and found beautiful results, including some which are comparable to some of ours, in [45, section 7] (see also [48]). They derive diagonalizability of states from a very different set of postulates.

5.6. Information-theoretic entropies and their relation to physics

So far we have considered entropy from a thermodynamic perspective. But entropies also arise in information theory, and as the GPT framework is mostly studied in quantum information theory, indeed there have been many results on entropy from a information-theoretic perspective. Our exposition will mainly follow [66], but has also been given in a slightly different formalism in [67].

Let $e = (e_1, \ldots, e_n)$ and $f = (f_1, \ldots, f_m)$ be two measurements such that there exists a map $M : \{1, \ldots, n\} \to \{1, \ldots, m\}$ with

$$\sum_{\{j|M(j)=k\}} e_j = f_k \qquad (k = 1, \ldots, m).$$

If M is bijective, then the measurement f is simply a *re-labelling* of e. If there exists a k with $M(j) \neq k \quad \forall j$, then because of the normalisation of the e-measurement, $f_k = 0$, i.e. f_k corresponds to a trivial outcome that never happens. If M is not injective, then f is a *coarse-graining* of e (or vice versa, e a *refinement* of f) in the sense that f is obtained from e by collecting several outcomes of e and giving them a common new outcome label (and by possibly adding the 0-effect a few times), see figure 5. In this sense, we do not care about which of the e_j triggered the new effect.

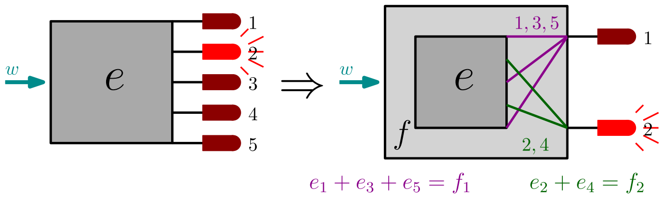

Figure 5. A coarse-graining of a measurement is created by having several measurement results trigger the same output.

However, there exist *trivial* refinements/coarse-grainings: for those, $e_j \propto f_{M(j)} \quad \forall j$. We write $e_j = p_j f_{M(j)}$. Then such a measurement can be obtained by performing f, and if outcome k is triggered, we activate a classical random number generator which generates the final outcome j among those j with $M(j) = k$ with probability

$$\frac{p_j}{\sum_{\{a|M(a)=k\}} p_a}.$$

Thus, a trivial refinement does not yield any additional information about the GPT-system. We call a measurement *fine-grained* if it does not have any non-trivial refinements. The set of fine-grained measurements on any state space A is denoted $\mathcal{E}_A^*$.

Now we consider the Rényi entropies [65], which are defined for probability distributions $\mathbf{p} = (p_1, \ldots, p_n)$ as

$$H_\alpha(\mathbf{p}) = \frac{1}{1-\alpha} \log\left(\sum_j p_j^\alpha\right),$$

where $\alpha \in (0, \infty) \setminus \{1\}$. Furthermore,

$$H_0(\mathbf{p}) \coloneqq \lim_{\alpha \to 0} H_\alpha(\mathbf{p}) = \log|\mathrm{supp}(\mathbf{p})|$$

with $\mathrm{supp}(\mathbf{p}) = \{p_j \mid p_j > 0\}$ is called the *max-entropy*, and

$$H_\infty(\mathbf{p}) \coloneqq \lim_{\alpha \to \infty} H_\alpha(\mathbf{p}) = -\log \max_j p_j$$

is called the *min-entropy*. Also,

$$H_1(\mathbf{p}) \coloneqq \lim_{\alpha \to 1} H_\alpha(\mathbf{p}) = -\sum_j p_j \log p_j = H(\mathbf{p})$$

is just the regular Shannon entropy H.

For $\alpha \in [0, \infty]$ and GPTs satisfying postulates 1 and 2, we generalise the classical Rényi entropies:

$$S_\alpha(\omega) \coloneqq H_\alpha(\mathbf{p}),$$

where $\omega = \sum_j p_j w_j$ is any decomposition into perfectly distinguishable pure states. According to theorem 3, the result is independent of the choice of decomposition. We have $S_1 = S$, the spectral entropy of definition 4.

Following [66], for every $\alpha \in [0, \infty]$ and $\omega \in \Omega_A$, we define the *order-α Rényi measurement entropy* as

$$\widehat{S}_\alpha(\omega) = \inf_{e \in \mathcal{E}_A^*} H_\alpha(e_1(\omega), e_2(\omega), \ldots),$$

where H_α on the right-hand side denotes the classical Rényi entropy. The *order-α Rényi decomposition entropy* is defined as

$$\check{S}_\alpha(\omega) \coloneqq \inf_{\omega = \sum_j q_j \varphi_j} H_\alpha(\mathbf{q}), \tag{11}$$

where the infimum is over all convex decompositions of ω into *pure* states $\varphi_j \in \Omega_A$.

The idea of measurement entropy is to characterise the state before a measurement. For example, in quantum theory, particles prepared in a state $|\psi\rangle$ which all give the same result in energy measurements would be said to be in an energy eigenstate. If instead we performed a position measurement, the resulting distribution of positions would have non-zero entropy. However, this entropy would arguably not come from the initial state, but from the measurement process itself due to the uncertainty principle.

Suppose we would like to prepare a state ω by using states of maximal knowledge (i.e. pure states) φ_j, and a random number generator which gives output j with probability p_j. Then the decomposition entropy quantifies the smallest information content (entropy) of a random number generator that would be necessary to build such a device. For more detailed operational interpretations of measurement and decomposition entropy, in particular for $\alpha = 1$, see [66, 67] Note that in quantum theory, measurement, decomposition and spectral Rényi entropies all coincide, with the $\alpha = 1$ case giving von Neumann entropy, $S(\omega) = -\mathrm{tr}(\omega \log \omega)$.

Our first result is that the spectral and measurement definitions of the entropies agree:

Theorem 10. *Consider any state space A which satisfies postulates 1 and 2. Then the Rényi entropies S_α and the Rényi measurement entropies $\widehat{S}_\alpha$ coincide, and upper-bound the Rényi decomposition entropy $\check{S}_\alpha$, i.e.*

$$\check{S}_\alpha(\omega) \leqslant S_\alpha(\omega) = \widehat{S}_\alpha(\omega) \text{ for all } \omega \in \Omega_A,\ \alpha \in [0, \infty].$$

In particular, for $\alpha = 1$, the measurement entropy $\widehat{S}$ is the same as the spectral entropy S from definition 4, which we have identified with thermodynamical entropy H in observation 5.

The inequality $\check{S}_\alpha \leqslant S_\alpha$ is easy to see: for a decomposition $\omega = \sum_i p_i \omega_i$ into perfectly distinguishable pure states ω_i, the states ω_i can also be seen as a fine-grained measurement, yielding outcome probabilities P_i. So taking the infimum over all decompositions gives *at most* $H_\alpha(\mathbf{p}) = S_\alpha(\omega)$. The equality between S_α and $\widehat{S}_\alpha$ is shown in the appendix.

We do not know in general whether postulates 1 and 2 imply that $\check{S}_\alpha = S_\alpha$ for all α. Interestingly, we know it for $\alpha = 2$ and $\alpha = \infty$:

Theorem 11. *If a state space satisfies postulates 1 and 2, then $\check{S}_2(\omega) = S_2(\omega)$ and $\check{S}_\infty(\omega) = S_\infty(\omega)$ for all states ω.*

Proof. To give the reader an idea of the kind of arguments involved, we present the proof for S_2, but defer the proof for S_∞ to the appendix. If $\omega = \sum_j p_j \omega_j$ is any convex decomposition into a maximal set of perfectly distinguishable pure states (without loss of generality $p_1 \geqslant p_2 \geqslant \ldots$), and $\omega = \sum_j q_j \varphi_j$ any (other) convex decomposition into pure states φ_j (also with $q_1 \geqslant q_2 \geqslant \ldots$,) then
$\sum_j p_j^2 = \langle \omega, \omega \rangle = \sum_j q_j^2 + \sum_{j \neq k} q_j q_k \langle \varphi_j, \varphi_k \rangle \geqslant \sum_j q_j^2$ since $\langle \varphi_j, \varphi_k \rangle \geqslant 0$. Thus, we have

$$S_2(\omega) = -\log \sum_j p_j^2 \leqslant -\log \sum_j q_j^2 = H_2(\mathbf{q}),$$

and since $\check{S}_2(\omega)$ is defined as the infimum over the right-hand side, we obtain that $\check{S}_2(\omega) \geqslant S_2(\omega)$; we find the converse inequality in theorem 10. □

We do not know whether the same identity holds for the most interesting case $\alpha = 1$, the case of standard thermodynamic entropy $S = S_1$. In the max-entropy case $\alpha = 0$, however, we have a surprising relation to higher-order interference:

Theorem 12. *Consider a state space satisfying postulates 1 and 2. Then the following statements are all equivalent:*

(i) *The state space does not have third-order interference.*

(ii) *The measurement and decomposition versions of max-entropy coincide, i.e. $\check{S}_0(\omega) = S_0(\omega)$ for all states ω.*

(iii) *The state space is either classical, or one on the list of theorem 2.*

(iv) *If ω is a pure state and P_F any orthogonal projection onto any face F, then $P_F\omega$ is a multiple of a pure state.*

(v) *The 'atomic covering property' of quantum logic holds.*

The equivalences (i) ⇔ (iii) ⇔ (iv) ⇔ (v) are shown in [43]; our new result is the equivalence to (ii), which is shown in the appendix.

Absence of third-order interference is meant in the sense of equation (5), as introduced originally by Sorkin [76]: only pairs of mutually exclusive alternatives can possibly interfere. It is interesting that this is related to an information-theoretic property of max-entropy S_0, as given in (ii). We do not currently know whether S_0 (or, in particular, the identity of $\check{S}_0$ and S_0) has any thermodynamic relevance in the class of theories that we are considering, but it certainly does within quantum theory, where it attains operational meaning in single-shot thermodynamics [28, 29].

As (iii) shows, this theorem is closely related to open problem 1: it gives properties of conceivable state spaces that satisfy postulates 1 and 2, but are not on the list of known examples (namely, they do not satisfy any of (i)–(v)). Similarly, (iv) shows the relation of higher-order interference to open problem 2, and (v) relates all these items to quantum logic. In fact, one can show that postulates 1 and 2 imply that the set of faces of the state space has the structure of an *orthomodular lattice*, which is often seen as the definition of quantum logic. For readers who are familiar with the terminology of quantum logic, we give some additional remarks in section A.3 in the appendix.

6. Conclusions

As discussed in the introduction, many works (dating back at least to the 1950s) have considered quantum theory as just one particular example of a probabilistic theory: a single point in a large space of theories that contains classical probability theory, as well as many other possibilities that are non-quantum and non-classical. More recent works have focused on the information-theoretic properties of quantum theory, for example deriving quantum theory as the unique structure that satisfies a number of information-theoretic postulates.

Rather than attempt a derivation of quantum theory from postulates, this paper has examined the thermodynamic properties of quantum theory and of those theories that are similar enough to quantum theory to admit a good definition of thermodynamic entropy, and of some version of the second law. Postulate 1 states that there is a reversible transformation between any two sets of n distinguishable pure states. This can be thought of as an expression of the universality of the representation of information, in particular that a choice of basis is arbitrary, and also allows for reversible microscopic dynamics, as is crucial for thermodynamics. Postulate 2 states that every state can be written as a convex mixture of perfectly distinguishable pure states. This ensures that a mixed state describing an ensemble of many particles can be treated as if each particle has an unknown microstate, drawn from a set of distinguishable possibilities.

Much follows from postulates 1 and 2, without needing to assume any other aspects of the standard formalism of quantum theory. In order to derive thermodynamic conclusions, we considered the argument originally employed by von Neumann in his derivation of the mathematical expression for the thermodynamic entropy of a quantum state. The argument involves a thought experiment with a gas of quantum particles in a box, and semi-permeable membranes that allow a particle to pass or not depending on the outcome of a quantum measurement. By applying the same thought experiment, we showed that given any theory satisfying postulates 1 and 2, there is a unique expression for the the thermodynamic entropy, equal to both the spectral entropy and the measurement entropy. By way of contrast, a fictitious system defined by a square state space, which arises as Alice's local system of an entangled pair producing stronger-than-quantum 'PR box' correlations, does not satisfy either Postulate. This system—the gbit—does not admit a sensible notion of thermodynamic entropy, at least not one that is given to it by the von Neumann or Petz arguments. While many works have discussed the inability of quantum theory to produce arbitrarily strong nonlocal correlations, this connection with thermodynamics deserves further investigation. It would be very interesting, for example, if Tsirelson's bound on the strength of quantum nonlocal correlations could be derived from a thermodynamic argument.

There are many other consequences of postulates 1 and 2 for both thermodynamic and information-theoretic entropies. For example, a form of the second law holds in that neither projective measurements nor mixing procedures can decrease the thermodynamic entropy. The spectral and measurement order-α Renyi entropies coincide for any α. The spectral and decomposition order-α Renyi entropies coincide for $\alpha = 2$ or ∞. An open question is whether any theory satisfying postulates 1 and 2 is completely satisfactory from the thermodynamic point of view. While the von Neumann and Petz arguments can be run with no trouble in the presence of postulates 1 and 2 as we have shown, there could still be a different physical scenario, in which theories would fail to exhibit sensible behaviour unless they have even more of the structure of quantum theory.

Finally, another major open question is whether quantum-like theories exist, satisfying postulates 1 and 2, that are distinct from quantum theory in that they admit higher-order interference. Roughly speaking, this means that three or more possibilities can interfere in order to produce an overall amplitude, unlike in quantum theory, where different possibilities only interfere in pairs. We extend the results of [43], where it was shown that in the context of postulates 1 and 2 the existence of higher-order interference is equivalent to each of three other statements. We provide an equivalent entropic condition: there is higher-order interference if and only if the measurement and decomposition versions of the max entropy do not coincide.

Our understanding of quantum theory would be greatly improved if higher-order interference could be ruled out by simple information-theoretic, thermodynamic, or other physical arguments. On the other hand, if theories with higher-order interference exist and are eminently sensible, an immediate question is whether an experimental test could be performed to distinguish such a theory from quantum theory. While previous experiments [97–102] only tested for a zero versus non-zero value of higher-order interference, sensible higher-order theories that satisfy postulates 1 and 2 (if they exist) could help to inform future experiments by supplying concrete models that can be tested against standard quantum theory.

Acknowledgments

We would like to thank Matt Leifer for many useful discussions, and we are grateful to the participants of the 'Foundations of Physics working group' at Western University for helpful feedback. We would also like to thank

Giulio Chiribella and Carlo Maria Scandolo for coordinating the arXiv posting of their work with us. This research was supported in part by Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is supported by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Research, Innovation and Science. This research was undertaken, in part, thanks to funding from the Canada Research Chairs programme. This research was supported by the FQXi Large Grant 'Thermodynamic versus information theoretic entropies in probabilistic theories'. HB thanks the Riemann Center for Geometry and Physics at the Institute for Theoretical Physics, Leibniz University Hannover, for support as a visiting fellow during part of the time this paper was in preparation.

Appendix

A.1. Proofs

A.1.1. Proof that observables are well-defined. In this appendix, a decomposition of a state into perfectly distinguishable pure states (which always exists due to postulate 2) will be called a 'classical decomposition'.

Lemma 13. *Assume postulates 1 and 2. Let $F \neq \{0\}$ be a face of A_+ and $\omega \in \Omega_A \cap F$. Then there exists a classical decomposition $\omega = \sum_j p_j \omega_j$ with $\omega_j \in F$ for all j.*

Proof. Let $\omega = \sum_j p_j \omega_j$ be a classical decomposition with $p_j \neq 0$. As $\omega \in F$ and F a face, $\omega_j \in F$ for all j. □

Proof of theorem 3. Let $x \in A$ be arbitrary. By lemma 5.46 from [62] there exists a frame $\{\omega_j\}$ and $x_j' \in \mathbb{R}$ such that $x = \sum_j x_j' \omega_j$. We extend $\{\omega_j\}$ to a maximal frame by adding $x_j' := 0$ for the new indices j. Now we group together the j with the same x_j' value, and by relabelling we find that $x = \sum_{k=1}^n x_k \sum_i \omega_{k;i}$ where the x_k are pairwise different values of the x_j' and the $\omega_{k;i}$ are the ω_j that belong to this x_j' value. For any given k, the $\omega_{k;i}$ generate a face F_k with projective unit $u_k = \sum_i \omega_{k;i}$.

Therefore we find a decomposition $x = \sum_{k=1}^n x_k u_k$ with x_k pairwise different real numbers and u_k order units of faces F_k and $\sum_{k=1}^n u_k = u_A$.

Now we show that the faces F_k are mutually orthogonal:

Let $\omega \in F_k$ be an arbitrary normalised state. By lemma 13 it has a classical decomposition $\omega = \sum_j p_j \omega_j^{(k)}$ which uses only pure states $\omega_j^{(k)} \in F_k$. W.l.o.g. we assume that these pure states form a generating frame of F_k, by extending the frame and adding $p_j = 0$ to the decomposition. Consider another face F_m, i.e. $m \neq k$. Likewise to ω, let $\omega' \in F_m$ be an arbitrary normalised state and $\omega' = \sum_j q_j \omega_j^{(m)}$ be a classical decomposition with $\omega_j^{(m)}$ a generating frame for F_m. For the other faces define $\omega_j^{(i)} := \omega_{i;j}$. Then $u_i = \sum_j \omega_j^{(i)}$ and in total $u_A = \sum_i u_i = \sum_i \sum_j \omega_j^{(i)}$. As $\langle \nu, \nu \rangle = 1$ for all pure states $\nu \in \Omega_A$, this implies that the $\omega_j^{(i)}$ are mutually orthogonal:

$$1 = u_A(\omega_h^{(g)}) = \sum_i \sum_j \langle \omega_j^{(i)}, \omega_h^{(g)} \rangle = 1 + \sum_{(i,j) \neq (g,h)} \langle \omega_j^{(i)}, \omega_h^{(g)} \rangle$$

and therefore $\langle \omega_j^{(i)}, \omega_h^{(g)} \rangle \geqslant 0$ implies $\langle \omega_j^{(i)}, \omega_h^{(g)} \rangle = 0$ for all $(i, j) \neq (g, h)$. Thus we find $\langle \omega, \omega' \rangle = \sum_j \sum_b p_j q_b \langle \omega_j^{(k)}, \omega_b^{(m)} \rangle = 0$ because $m \neq k$. As $\omega \in F_k$ and $\omega' \in F_m$ were arbitrary (normalised) states, this implies that F_k and F_m are orthogonal. As $k \neq m$ were arbitrary, all the faces are mutually orthogonal.

Now we will show that the decomposition $x = \sum_j x_j u_j$ is unique. So assume there are two decompositions $x = \sum_{j=1}^{n_a} a_j u_j^{(a)} = \sum_{j=1}^{n_b} b_j u_j^{(b)}$ with $a_j \in \mathbb{R}$ pairwise different and projective units $u_j^{(a)}$ that add up to the order unit (analogously for b) and belong to pairwise orthogonal faces $F_j^{(a)}$. W.l.o.g. we assume that the a_j and b_j are ordered by size, i.e. $a_1 < a_2 < \ldots < a_{n_a}$. We want to show $a_1 = b_1$. The $u_j^{(a)}$ generate the faces $F_j^{(a)}$. Let $\omega_{j;i}^{(a)}$ be a generating frame for the face $F_j^{(a)}$, especially $\sum_i \omega_{j;i}^{(a)} = u_j^{(a)}$. As the faces are mutually orthogonal and the projective units add up to u_A, the $\omega_{j;i}^{(a)}$ form a maximal frame; in particular they add up to u_A (likewise for b). Therefore:

$$a_1 = \langle \omega_{1;j}^{(a)}, x \rangle = \sum_{k,i} b_k \langle \omega_{1;j}^{(a)}, \omega_{k;i}^{(b)} \rangle \geqslant \sum_{k,i} b_1 \langle \omega_{1;j}^{(a)}, \omega_{k;i}^{(b)} \rangle = b_1 u_A(\omega_{1;j}^{(a)}) = b_1.$$

Analogously show $b_1 \geqslant a_1$, i.e. $b_1 = a_1$ in total.

Now suppose there was a $k > 1$ and an i with $\langle \omega_{1;j}^{(a)}, \omega_{k;i}^{(b)} \rangle \neq 0$, i.e. $\langle \omega_{1;j}^{(a)}, \omega_{k;i}^{(b)} \rangle > 0$. Then

$$a_1 = \langle \omega_{1;j}^{(a)}, x \rangle = \sum_{k,i} b_k \langle \omega_{1;j}^{(a)}, \omega_{k;i}^{(b)} \rangle = a_1 \sum_i \langle \omega_{1;j}^{(a)}, \omega_{1;i}^{(b)} \rangle + \sum_{k>1,i} b_k \langle \omega_{1;j}^{(a)}, \omega_{k;i}^{(b)} \rangle > a_1 \sum_i \langle \omega_{1;j}^{(a)}, \omega_{1;i}^{(b)} \rangle$$
$$+ \sum_{k>1,i} a_1 \langle \omega_{1;j}^{(a)}, \omega_{k;i}^{(b)} \rangle = a_1 \sum_{k,i} \langle \omega_{1;j}^{(a)}, \omega_{k;i}^{(b)} \rangle = a_1 u_A(\omega_{1;j}^{(a)}) = a_1.$$

This is a contradiction. Thus $\langle \omega_{1;j}^{(a)}, \omega_{k;i}^{(b)} \rangle = 0$ for all $k > 1$ and i. Therefore we find $u_1^{(b)}(\omega_{1;i}^{(a)}) = \sum_j \langle \omega_{1;j}^{(b)}, \omega_{1;i}^{(a)} \rangle = \sum_{j,k} \langle \omega_{k,j}^{(b)}, \omega_{1;i}^{(a)} \rangle = u_A(\omega_{1;i}^{(a)}) = 1$ and analogously $u_1^{(a)}(\omega_{1;i}^{(b)}) = 1$. By proposition 5.29 from [62], we have $\Omega_A \cap F = \{\omega \in \Omega_A \mid u_F(\omega) = 1\}$. Therefore a generating frame of $F_1^{(a)}$ is contained in $F_1^{(b)}$ and vice versa. Thus we find $F_1^{(a)} = F_1^{(b)}$ and $u_1^{(a)} = u_1^{(b)}$.

For the remaining indices, we construct an inductive proof: choose $L \in \mathbb{R}$ large enough such that $a_1 + L > \max\{a_{n_a}, b_{n_b}\}$, and define $x' := x + L \cdot u_1^{(a)}$, i.e. $x' = \sum_{j=1}^{n_a} (a_j + \delta_{j,1} \cdot L) u_j^{(a)} = \sum_{j=1}^{n_b} (b_j + \delta_{j,1} \cdot L) u_j^{(b)}$. Furthermore defining $a_1' := a_2, a_2' := a_3, \ldots, a_{n_a}' := a_1 + L, u_1^{(a')} := u_2^{(a)}, u_2^{(a')} := u_3^{(a)}, \ldots, u_{n_a}^{(a')} := u_1^{(a)}$ and likewise for b_j', we find $x' = \sum_{j=1}^{n_a} a_j' u_j^{(a')} = \sum_{j=1}^{n_b} b_j' u_j^{(b')}$ with $a_1' < a_2' < \ldots < a_{n_a}'$ and $b_1' < b_2' < \ldots < b_{n_b}'$. Repeating the exact same procedure as before, we obtain $a_1' = b_1'$ and $u_1^{(a')} = u_1^{(b')}$, i.e. $a_2 = b_2$ and $u_2^{(a)} = u_2^{(b)}$. We iterate to find $a_j = b_j$ and $u_j^{(a)} = u_j^{(b)}$ for all j. Note that as all maximal frames have the same size and as the projective units add up to u_A, necessarily $n_a = n_b$.

At last we construct the projective measurement that corresponds to measuring the observable x: for F_k, let P_k be the orthogonal projector onto the span of F_k (in particular, $P_k : A \to \mathrm{span}(F_k)$ surjective). We know that these projectors are positive and linear and satisfy $u_A \circ P_k = u_k$. Furthermore $0 \leqslant u_k = u_A \circ P_k \leqslant u_A$ and $\sum_k u_A \circ P_k = \sum_k u_k = u_A$, i.e. we obtain a well-defined measurement; therefore the P_k form a well-defined instrument. As they are projectors, the P_k leave the elements of F_k unchanged. □

A.1.2. Proof of observation 5. In order to show that $H(\omega) = S(\omega)$ is consistent with assumptions 1, we only have to show that $\omega \mapsto S(\omega)$ is continuous, to comply with assumption (d). According to theorem 10 (which we will prove below), the spectral entropy $S(\omega)$ equals measurement entropy $\widehat{S}(\omega)$. But it is well-known [67] and easy to see from its definition that $\widehat{S}$ is continuous.

It remains to show equation (9). So let $\omega = \sum_j p_j \omega_j$ be any decomposition of ω into perfectly distinguishable, not necessarily pure states ω_i. Decompose all the ω_i into perfectly distinguishable pure states $\omega_j^{(i)}$, i.e. $\omega_j = \sum_i q_j^{(i)} \omega_j^{(i)}$. Perfectly distinguishable states live in orthogonal faces, thus $\langle \omega_i, \omega_j \rangle = 0$ for $i \neq j$ (note that this is a conclusion that follows from postulates 1 and 2, but could not be drawn from bit symmetry alone in [64]). Thus, we also have $\langle \omega_i^{(j)}, \omega_k^{(l)} \rangle = 0$ for $i \neq k$ or $j \neq l$, and so $\omega = \sum_{ij} p_j q_j^{(i)} \omega_j^{(i)}$ is a decomposition of ω into perfectly distinguishable pure states. Define the real function η: $[0, 1] \to \mathbb{R}$ via $\eta(x) := -x \log x$ for $x > 0$ and $\eta(0) = 0$. Due to theorem 3 and $\eta(xy) = -xy \log x - xy \log y$, we have

$$\eta(\omega) = \sum_{ij} \eta(p_j q_j^{(i)}) \omega_j^{(i)} = -\sum_{ij} p_j q_j^{(i)} \log p_j \omega_j^{(i)} - \sum_{ij} p_j q_j^{(i)} (\log q_j^{(i)}) \omega_j^{(i)},$$

and therefore

$$S(\omega) = u_A(\eta(\omega)) = -\sum_{ij} p_j q_j^{(i)} \log p_j - \sum_{ij} p_j q_j^{(i)} \log q_j^{(i)} = -\sum_j p_j \log p_j + \sum_j p_j \underbrace{\left(-\sum_i q_j^{(i)} \log q_j^{(i)}\right)}_{S(\omega_j)}$$

This completes the proof of observation 5. □

A.1.3. Proof of the second half of theorem 11. Use the notation of the first half of the proof. We claim that $\max_{\varphi \in \Omega} \langle \omega, \varphi \rangle = p_1$. The inequality '$\geqslant$' is trivial (consider the special case $\varphi = \omega_1$). To see the inequality '$\leqslant$', note that $\langle \omega, \varphi \rangle = \sum_j p_j \lambda_j$, where $\lambda_j := \langle \omega_j, \varphi \rangle \in [0, 1]$ satisfies $\sum_j \lambda_j = \langle \sum_j \omega_j, \varphi \rangle = \langle u, \varphi \rangle = 1$, and so $\langle \omega, \varphi \rangle \leqslant p_1$ for all φ. Thus

$$S_\infty(\omega) = -\log p_1 = -\log \max_{\varphi \in \Omega} \langle \omega, \varphi \rangle \leqslant -\log \langle \omega, \varphi_1 \rangle = -\log \left(q_1 + \sum_{j \geqslant 2} q_j \langle \varphi_j, \varphi_1 \rangle \right) \leqslant -\log q_1 = H_\infty(\mathbf{q}).$$

Similarly as in the first part of the proof, we obtain $\check{S}_\infty(\omega) \geqslant S_\infty(\omega)$. The converse inequality from theorem 10 for $\alpha = \infty$ concludes the proof. □

A.1.4. Proof of Klein's inequality and the second law for projective measurements. We consider an ensemble of systems described by an arbitrary state $\omega \in \Omega_A$. To all systems of this ensemble we apply a projective measurement described by orthogonal projectors P_a which form an instrument, resulting in a new ensemble state ω'. The P_a project onto the linear span of faces F_a that replace the eigenspaces from quantum theory. We want to show that the measurement cannot decrease the entropy of the ensemble, i.e.

$$S(\omega') \geqslant S(\omega).$$

We decompose the proof into several steps. Our basic idea follows the proof of a similar statement for quantum theory in [50]: we reduce the proof of the second law to Klein's inequality. But as we do not have access to an underlying pure state Hilbert space, we will need to use a different argument for why Klein's inequality implies the second law for projective measurements.

So at first we prove Klein's inequality, adapting the proof of [50]. We note that a similar proof has also been found by Scandolo [48], albeit under different assumptions.

Proof of theorem 9. We consider two arbitrary states ω, ν with classical decompositions $\omega = \sum_j p_j \omega_j$, $\nu = \sum_k q_k \nu_k$, where w.l.o.g. the ω_j and the ν_k form maximal frames. We define the matrix $P_{jk} := \langle \omega_j, \nu_k \rangle$. All its components are non-negative, i.e. $P_{jk} \geqslant 0$, because the scalar product itself is non-negative for all states. As all maximal frames have the same size, the matrix is a square matrix; as maximal frames sum to u_A, the rows and columns sum to one: $\sum_j P_{jk} = \sum_k P_{jk} = 1$. Thus, we get

$$S(\omega\|\nu) = -S(\omega) - \langle \omega, \log \nu \rangle = \sum_j p_j \log p_j - \sum_{jk} p_j \log q_k \langle \omega_j, \nu_k \rangle = \sum_j p_j \left(\log p_j - \sum_k P_{jk} \log q_k \right).$$

We define $r_j := \sum_k P_{jk} q_k$. Note that the r_j form a probability distribution: $r_j \geqslant 0$ and $\sum_j r_j = \sum_k \sum_j P_{jk} q_k = \sum_k q_k = 1$. Using the strict concavity of the logarithm, we find:

$$\log r_j = \log \left(\sum_k P_{jk} q_k \right) \geqslant \sum_k P_{jk} \log q_k.$$

Therefore we get

$$S(\omega\|\nu) = \sum_j p_j \left(\log p_j - \sum_k P_{jk} \log q_k \right) \geqslant \sum_j p_j (\log p_j - \log r_j) = \sum_j p_j \log \left(\frac{p_j}{r_j} \right).$$

We recognise the last expression as the classical relative entropy of the probability distributions p_j and r_j. This classical relative entropy has the important property that it is never negative. Thus:

$$S(\omega\|\nu) \geqslant 0.$$

□

In order to get the main proof less convoluted, we will state some technical parts as lemmas.

Lemma 14. *Assume postulate 1 and 2. Consider orthogonal projectors P_j which form an instrument. Then the P_j are mutually orthogonal:*

$$P_k P_j = \delta_{jk} P_j.$$

Proof. We prove $P_k P_j \omega = 0$ for all $\omega \in A$, $j \neq k$. If $P_j \omega = 0$ this is trivial, so from now on assume $P_j \omega \neq 0$. As the cone is generating (i.e. $\mathrm{Span}(A_+) = A$) and the projectors linear, it is sufficient to show $P_k P_j \omega = 0$ for all $w \in A_+$. As P_j is positive, $P_j \omega \neq 0$ implies that $(u_A \circ P_j)(\omega) > 0$ because only the zero-state is normalised to 0. Using $u_A = u_A \circ (\sum_j P_j) = \sum_j u_A \circ P_j$ and $P_j P_j = P_j$:

$$1 = u_A \left(\frac{P_j \omega}{(u_A \circ P_j)(\omega)} \right) = \left(u_A \circ \sum_k P_k \right) \left(\frac{P_j \omega}{(u_A \circ P_j)(\omega)} \right) = u_A \left(\frac{P_j P_j \omega}{(u_A \circ P_j)(\omega)} \right) + u_A \left(\frac{\sum_{k|k\neq j} P_k P_j \omega}{(u_A \circ P_j)(\omega)} \right)$$
$$= u_A \left(\frac{P_j \omega}{(u_A \circ P_j)(\omega)} \right) + u_A \left(\frac{\sum_{k|k\neq j} P_k P_j \omega}{(u_A \circ P_j)(\omega)} \right) \quad \Rightarrow \quad 0 = u_A \left(\frac{\sum_{k|k\neq j} P_k P_j \omega}{(u_A \circ P_j)(\omega)} \right) = \sum_{k|k\neq j} u_A (P_k P_j \omega).$$

As the projectors are positive and only the zero-state is normalised to 0, this shows $P_k P_j \omega = 0$ for $k \neq j$. □

Lemma 15. *Assume postulates 1 and 2. Consider an orthogonal projector P which projects onto the linear span of a face F of A_+. Then for all states $\omega \in A_+$ we find $P\omega \in F$.*

Proof. From basic convex geometry (see e.g. proposition 2.10 in [63]), we know that $F = \mathrm{span}(F) \cap A_+$. Since P is positive, we have $P\omega \in A_+$; furthermore, since P projects onto F, we have $P\omega \in \mathrm{span}(F)$, thus $P\omega \in F$. □

Proof of theorem 6. We know that $S(\omega\|\omega') = -S(\omega) - \langle\omega, \log\omega'\rangle \geqslant 0$. As in theorem 11.9 from [50], we claim $-\langle\omega, \log\omega'\rangle = S(\omega')$ and therefore $-S(\omega) + S(\omega') \geqslant 0$. Thus we only have to prove $-\langle\omega, \log\omega'\rangle = S(\omega')$. But as we do not have access to an underlying pure state Hilbert space, our proof is different from [50].

By lemma 14, the P_a are mutually orthogonal, i.e. $P_a P_b = \delta_{ab} P_b$. By symmetry of the P_a also the $P_a\omega$ are mutually orthogonal: $\langle P_a\omega, P_b\omega\rangle = \langle\omega, P_a P_b\omega\rangle = 0$ for $a \neq b$. This also shows that the F_a are mutually orthogonal. If $P_a\omega = 0$ we use the decomposition $P_a\omega = u_A(P_a\omega)\sum_k r_{ak} w_{ak}$ with $r_{ak} = \delta_{ak}$ and w_{ak} an arbitrary generating frame of F_a. If $P_a\omega \neq 0$, then $\frac{P_a\omega}{u_A(P_a\omega)} \in F_a \cap \Omega_A$ and by lemma 13, there is a classical decomposition $\frac{P_a\omega}{u_A(P_a\omega)} = \sum_k r_{ak}\omega_{ak}$ with $\omega_{ak} \in F_a$. We complete the ω_{ak} to generating frames of the F_a by adding terms with $r_{ak} = 0$. As we are using classical decompositions/frames, we know $\langle\omega_{aj}, \omega_{ak}\rangle = \delta_{jk}$. Furthermore, as the F_a are mutually orthogonal, we know $\langle\omega_{aj}, \omega_{bk}\rangle = 0$ for $a \neq b$.

We note that the the w_{aj} form a maximal frame:

$$u_A = \sum_a u_A \circ P_a = \sum_a u_{F_a} = \sum_a \sum_j \omega_{aj}.$$

For $a \neq b$ we have $P_b\omega_{aj} = P_b P_a\omega_{aj} = 0$, so we have a classical decomposition

$$\omega' = \sum_a P_a\omega = \sum_a \sum_j u_A(P_a\omega) r_{aj}\omega_{aj}$$

with ω_{aj} a maximal frame that satisfies $P_a\omega_{bj} = \delta_{ab}\omega_{bj}$. Note that we do not need to normalise ω' as the measurement itself is required to be normalised. Using

$$\sum_a P_a \log\omega' = \sum_{bj} \log(u_A(P_b\omega) r_{bj}) \sum_a P_a\omega_{bj} = \sum_{bj} \log(u_A(P_b\omega) r_{bj})\omega_{bj} = \log\omega'$$

and

$$-S(\omega') = -\sum_{bj} (u_A(P_b\omega) r_{bj}) \log(u_A(P_b\omega) r_{bj}) = \langle\omega', \log\omega'\rangle \tag{12}$$

as well as the symmetry of the P_a we finally find:

$$-S(\omega') = \langle\omega', \log\omega'\rangle = \left\langle \sum_a P_a\omega, \log\omega' \right\rangle = \left\langle \omega, \sum_a P_a \log\omega' \right\rangle = \langle\omega, \log\omega'\rangle. \tag{13}$$

□

A.1.5. Proof that measurement and spectral entropies are identical. In the main text we encountered different ways to define the entropy. One of them is to adapt classical entropy definitions by using the coefficients of a classical decomposition. Another is to adapt classical entropy definitions by using measurement probabilities and minimising over all fine-grained measurements. Here we will show that in the context of postulates 1 and 2, these two concepts yield the same Rényi entropies.

To prove this, we will first analyse fine-grained measurements in further detail. The results will allow us to reproduce the quantum proof found in [66] for our GPTs.

Lemma 16. *Assume postulates 1 and 2. Consider an arbitrary fine-grained measurement $(e_1, \ldots, e_n)$. Then for all j there exist some $c_j \in [0, 1]$ and a pure state $\omega_j \in \Omega_A$ such that $e_j = c_j\langle\omega_j, \cdot\rangle$.*

Proof. If $e_j = 0$, we can just take $c_j = 0$ and any pure state ω_j. So from now on assume $e_j \neq 0$.

Because of self-duality there exists some $\omega' \in A_+$ such that $\langle\omega', \cdot\rangle = e_j$. As $e_j \neq 0$ also $\omega' \neq 0$ and therefore $u_A(\omega') \neq 0$. With $A_+ = \mathbb{R}_{\geqslant 0} \cdot \Omega_A$ and $c_j := u_A(\omega') > 0$ there exists an $\omega \in \Omega_A$ such that $\omega' = c_j \cdot \omega$. We want to prove that ω is pure, so assume it was not pure. Then it has a classical decomposition $\omega = \sum_{k=0}^N p_k\omega_k$ with $p_k > 0$ and $N \geqslant 1$. By relabelling we can assume $j = n$, i.e. we consider $e_n = c_j\sum_{k=0}^N p_k\langle\omega_k, \cdot\rangle$. Define a measurement $(e_1', \ldots, e_{n+N}')$ by $e_k' := e_k$ for all $k = 1, 2, \ldots, n-1$ and $e_{n+i}' := c_j p_i\langle\omega_i, \cdot\rangle$ for all $i = 0, 1, \ldots, N$. Because of $0 \leqslant c_j p_i\langle\omega_i, \cdot\rangle = e_{n+i}'$ and $\sum_{k=1}^{n+N} e_k' = \sum_{k=1}^{n-1} e_k + \sum_{i=0}^N c_j p_i\langle\omega_i, \cdot\rangle = \sum_{k=1}^n e_k = u_A$ this is a well-defined measurement.

Now define M: $\{1, \ldots, n+N\} \to \{1, \ldots, n\}$ by $M(i) := i$ for all $i = 1, \ldots, n-1$ and $M(i) := n$ for all $i = n, \ldots, n+N$. Then we get

$$\sum_{\{a|M(a)=i\}} e_a' = e_i \qquad \text{for } i < n \quad \sum_{\{a|M(a)=i\}} e_a' = \sum_{a=n}^{n+N} e_i' = e_n \qquad \text{for } i = n.$$

Thus the measurement $(e_1', \dots, e_{n+N}')$ is a refinement of $(e_1, \dots, e_n)$. With $e_n'(\omega_0) = c_j p_0 = e_n(\omega_0)$ and $e_n'(\omega_1) = 0 \neq e_n(\omega_1)$ we find that e_n' is not proportional to e_n, thus the fine-graining is non-trivial. This is in contradiction to our assumptions. Thus ω has to be pure. Furthermore $1 = u_A(\omega) \geqslant e_j(\omega) = c_j \langle \omega, \omega \rangle = c_j$.

So in total we have found that $e_j = c_j \langle \omega, \cdot \rangle$ with $\omega \in \Omega_A$ pure and $c_j \in [0, 1]$. □

Lemma 17. *Assume postulates 1 and 2. Let $\omega \in \Omega_A$ and $\omega = \sum_{j=1}^{d} p_j \omega_j$ be a decomposition into a maximal frame. Then the measurement that perfectly distinguishes the ω_j (i.e. $e_k(\omega_j) = \delta_{jk}$) can be chosen to be fine-grained.*

Proof. Define $e_j := \langle \omega_j, \cdot \rangle$. As maximal frames add up to the order unit, this is a well-defined measurement and it satisfies $e_j(\omega_k) = \delta_{jk}$. It remains to show that this measurement is fine-grained.

Consider a fine-graining e_k' with $e_i = \sum_{\{j|M(j)=i\}} e_j'$. By self-duality, there exist $c_j \geqslant 0$ and $\omega_j' \in \Omega_A$ such that $e_j' = c_j \langle \omega_j', \cdot \rangle$ and therefore $\sum_{\{j|M(j)=k\}} c_j \omega_j' = \omega_k$. As $1 = u_A(\omega_k) = \sum_{\{j|M(j)=k\}} c_j u_A(\omega_j') = \sum_{\{j|M(j)=k\}} c_j$ we find that $\sum_{\{j|M(j)=k\}} c_j \omega_j' = \omega_k$ is a convex decomposition of a pure state. This requires $c_j = 0$ or $\omega_j' = \omega_k$. In both cases $e_j' = c_j \langle \omega_k, \cdot \rangle = c_j e_k$ holds true for all j with $M(j) = k$. Therefore, the fine-graining is trivial. □

Lemma 18. *Assume postulates 1 and 2. Consider a fine-grained measurement $\mathbf{e} = (e_1, \dots, e_N) \in \mathcal{E}^*$. Then the maximal number of perfectly distinguishable states d (often denoted as N_A) satisfies $d \leqslant N$.*

Furthermore, consider a state $\omega \in \Omega_A$ with classical decomposition $\omega = \sum_{j=1}^{d} p_j \omega_j$ into a maximal frame. Define the vector $\mathbf{q} := (e_j(\omega))_{1 \leqslant j \leqslant N}$ of outcome probabilities and the N-component vector $\mathbf{p} = (p_1, \dots, p_d, 0, \dots, 0) \in \mathbb{R}^N$. Then $\mathbf{q} \prec \mathbf{p}$, i.e. there exists a bistochastic $N \times N$-matrix M with $\mathbf{q} = M\mathbf{p}$.

Proof. By lemma 16 there exist $c_j \in [0, 1]$ and pure $\omega_j' \in \Omega_A$ such that $e_j = c_j \langle \omega_j', \cdot \rangle$. Define $q_l := e_l(\omega) = c_l \langle \omega_l', \omega \rangle$. Using $\sum_{j=1}^{N} e_j = u_A$ and $\sum_{j=1}^{d} \nu_j = u_A$ for an arbitrary maximal frame $(\nu_1, \dots, \nu_d)$ we find:

$$\sum_{j=1}^{N} c_j = \sum_{j=1}^{N} c_j u_A(\omega_j') = u_A\left(\sum_{j=1}^{N} c_j \omega_j'\right) = u_A(u_A) = u_A\left(\sum_{j=1}^{d} \nu_j\right) = \sum_{j=1}^{d} u_A(\nu_j) = \sum_{j=1}^{d} 1 = d.$$

As $c_j \in [0, 1]$, $\sum_{j=1}^{N} c_j = d$ shows $d \leqslant N$.

Set $q_{l|j} := e_l(\omega_j)$, introduce the N-component vector $\mathbf{p} := (p_1, \dots, p_d, 0, \dots, 0)$ and use that measurement effects and states of maximal frames add up to the order unit:

$$\sum_{j=1}^{d} q_{l|j} p_j = \sum_{j=1}^{d} e_l(p_j \omega_j) = e_l(\omega) = q_l, \quad \sum_{j=1}^{d} q_{l|j} = \sum_{j=1}^{d} e_l(\omega_j) = c_l \sum_{j=1}^{d} \langle \omega_l', \omega_j \rangle = c_l u_A(\omega_l') = c_l,$$
$$\sum_{l=1}^{N} q_{l|j} = \sum_{l=1}^{N} e_l(\omega_j) = u_A(\omega_j) = 1.$$

For $j \leqslant d$ we define $M_{l,j} := q_{l|j}$. If $d < N$ we also define $M_{l,j} := \frac{1 - c_l}{N - d}$ for $N \geqslant j > d$. M is an $N \times N$-matrix and it is bistochastic: first of all, $M_{l,j} \geqslant 0$ for all l, j. Furthermore:

$$\sum_{l=1}^{N} M_{l,j} = \sum_{l=1}^{N} q_{l|j} = 1 \qquad \text{for } j \leqslant d,$$
$$\sum_{l=1}^{N} M_{l,j} = \sum_{l=1}^{N} \frac{1 - c_l}{N - d} = \frac{N - d}{N - d} = 1 \qquad \text{for } j > d,$$
$$\sum_{j=1}^{N} M_{l,j} = \sum_{j=1}^{d} q_{l|j} + (N - d) \cdot \frac{1 - c_l}{N - d} = c_l + 1 - c_l = 1.$$

This bistochastic matrix maps $\mathbf{p}$ to $\mathbf{q}$, i.e. $M \cdot \mathbf{p} = \mathbf{q}$:

$$\sum_{j=1}^{N} M_{l,j} p_j = \sum_{j=1}^{d} q_{l|j} p_j = q_l.$$

□

Now we come to the proof of the theorem:

Proof of theorem 10. Consider an arbitrary fine-grained measurement $(e_1, \dots, e_N)$ and an arbitrary state $\omega \in \Omega_A$ with classical decomposition $\omega = \sum_{j=1}^{d} p_j \omega_j$ into a maximal frame. Define $q_l := e_l(\omega)$ and the N-component vector $\mathbf{p} = (p_1, \dots, p_d, 0, \dots, 0)$. Let M be the bistochastic matrix from lemma 18 with $\mathbf{q} = M \cdot \mathbf{p}$. By Birkhoff's theorem, it is a convex combination of permutation matrices, i.e. $M = \sum_{\sigma \in S_N} a_\sigma P_\sigma$ for a probability distribution a_σ and permutation matrices P_σ. W.l.o.g. we only consider the Shannon entropy; the proof for the Rényi entropies works exactly the same way. As the Shannon entropy is Schur-concave and invariant under permutations:

$$H(\mathbf{q}) \geqslant \sum_{\sigma \in S_N} a_\sigma H(P_\sigma \cdot \mathbf{p}) = \sum_{\sigma \in S_N} a_\sigma H(\mathbf{p}) = H(\mathbf{p}) = S(\omega).$$

Furthermore $H(\mathbf{p}) = -\sum_{j=1}^{d} p_j \log p_j = S(\omega)$ is the entropy of a measurement that perfectly distinguishes the ω_j, i.e. $e_j(\omega_k) = \delta_{jk}$. Because of lemma 17, such a measurement can be chosen to be finegrained. Therefore we find:

$$\widehat{H}(\omega) = \inf_{\mathbf{e} \in \mathcal{E}^*} H(\mathbf{e}(\omega)) = H(\mathbf{p}) = S(\omega).$$

□

A.1.6. Proof of theorem 12. As mentioned in the main text, the equivalences (i) $\Leftrightarrow$ (iii) $\Leftrightarrow$ (iv) $\Leftrightarrow$ (v) are shown in [43]. We will now prove the equivalence (ii) $\Leftrightarrow$ (v), which proves theorem 12. Taking into account theorem 10, and formulating the atomic covering property in the context of theories that satisfy postulates 1 and 2, it remains to show the equivalence of the following two statements:

(ii′) For all states $\omega \in \Omega_A$, we have $\check{S}_0(\omega) \geqslant S_0(\omega)$.

(v′) If F is any face of A_+, and ω is any pure state, then the smallest face G that contains both F and ω has rank $|G| \leqslant |F| + 1$. (Note that this is trivial if $\omega \perp F$.)

We will first prove that (ii′) $\Rightarrow$ (v′), which is equivalent to $\neg$(v′) $\Rightarrow \neg$ (ii′). So suppose that there exists some face F of A_+ and a pure state ω such that the face G generated by both has rank $|G| \geqslant |F| + 2$. Let $\omega_1, \ldots, \omega_{|F|}$ be a frame that generates the face F. Then F is also generated by $\nu := \frac{1}{|F|}\sum_{j=1}^{|F|} \omega_j$, i.e. the normalised projective unit of F. This is because every face containing ν also contains all the ω_j (and vice versa), and F is the smallest face with this property.

Now consider the state $\frac{1}{2}\omega + \frac{1}{2}\nu$. The smallest face that contains this state must be G. If this state had a decomposition into $|F| + 1$ or fewer perfectly distinguishable pure states, then these would also generate G, and so $|G| \leqslant |F| + 1$, in contradiction to our assumption. Thus any decomposition of $\frac{1}{2}\omega + \frac{1}{2}\nu$ into perfectly distinguishable pure states uses at least $|F| + 2$ states with non-zero coefficients, i.e. $S_0(\frac{1}{2}\omega + \frac{1}{2}\nu) \geqslant \log(|F| + 2)$. But $\frac{1}{2}\omega + \frac{1}{2}\nu = \frac{1}{2|F|}\sum_{j=1}^{|F|}\omega_j + \frac{1}{2}\omega$ is a convex decomposition into $|F| + 1$ pure states, thus

$$\check{S}_0\left(\frac{1}{2}\omega + \frac{1}{2}\nu\right) \leqslant \log(|F| + 1) < \log(|F| + 2) \leqslant S_0\left(\frac{1}{2}\omega + \frac{1}{2}\nu\right).$$

It remains to show that (v′) $\Rightarrow$ (ii′). So suppose that (v′) holds, but that there is a state $\omega \in \Omega_A$ with $\check{S}_0(\omega) < S_0(\omega)$; we will show that this leads to a contradiction. By definition of $\check{S}_0$, if this is the case, then there exist pure states $\omega_1, \ldots, \omega_n$ with $n = \exp(\check{S}_0(\omega))$ and $p_1, \ldots, p_n \geqslant 0, \sum_i p_i = 1$, such that $\omega = \sum_{i=1}^{n} p_i \omega_i$. Using property (v′), and recursively looking at the faces generated by ω_1, generated by ω_1, ω_2, generated by $\omega_1, \omega_2, \omega_3$ and so forth, shows that the rank of the face G generated by $\omega_1, \ldots, \omega_n$ can be at most n. Since $\omega \in G$, this shows that ω can be decomposed into n or fewer pure perfectly distinguishable states. Therefore $S_0(\omega) \leqslant \log n = \check{S}_0(\omega)$. □

A.3. Some additional remarks on theorem 12 and quantum logic

The GPT framework has a close relation to quantum logic. This is not surprising, since much of the terminology of GPTs has appeared much earlier, in work on quantum logic and beyond. The approach via convex sets of states and observables can be traced back to Mackey [18] (who immediately made connections to quantum logic), and was developed further through the 1960s and beyond. A partial list of references includes [14, 20, 88], the last two of which offer axiomatic characterisations of quantum theory. Interaction with the quantum logic tradition continued, with the orthomodularity of the lattice of faces of the state and/or effect spaces often providing a point of contact, especially in Ludwig's work [20]. Also closely related to the convex sets approach was the work of Foulis and Randall [10–13] who, for example, studied ways of combining probabilistic systems.

Since postulates 1 and 2 imply that the state cone A_+ is self-dual (so coincides with the effect cone), and that each face of this cone is the intersection of the cone with the image of a *filter* (equivalently given self-duality, a *compresssion* in the sense of [24]), we have from these postulates alone, via e.g. [24], theorem 8.10, that the face lattice is orthomodular. The notion of orthomodular partially ordered set, or its special case, the notion of orthomodular lattice, is often taken to *define* the notion of quantum logic, so we can say that postulates 1 and 2 imply that the face lattice of the cone of states, (equivalently of the cone of effects, or of the set of normalised states) is a quantum logic.

The covering property, in its most common variant the *atomic covering property*, states that for any element x of the lattice and any atom a not below or equal to x, $a \vee x$ covers x[17].The equivalence of (iv) and (v), in a setting more general than postulates 1 and 2, is proposition 9.7 of [24] (first appearing in proposition 4.2 in [25] and the discussion preceding it). Along with orthomodularity, the covering law was one of the assumptions of Piron's famous lattice-theoretic characterisation ([94]; also [17] and see the discussion in [27] or for more detail and proofs, [26 pp 18–38, 114–122]) of a class of lattices close to, although larger than, that of real, complex, and quaternionic quantum theory. A generalisation of the covering law was also used in Ludwig's axiomatization (see e.g. [20][18]) of quantum theory within the convex sets framework, in which the relevant lattice is a lattice of faces of the state space (equivalent to a lattice of extremal effects in his context), and the result characterised real, complex, and quaternionic quantum theory[19].

References

[1] Everett H 1973 The theory of the universal wave function *The Many-Worlds Hypothesis of Quantum Mechanics* (Princeton, NJ: Princeton University Press) pp 3–137
[2] Bohm D 1952 A suggested interpretation of the quantum theory in terms of hidden variables I & II *Phys. Rev.* **85** 166–93
[3] Fuchs C A 2010 QBism, the perimeter of quantum bayesianism arXiv:1003.5209
[4] Cabello A 2015 Interpretations of quantum theory: a map of madness arXiv:1509.04711
[5] Bell J S 1964 On the Einstein Podolsky Rosen paradox *Physics* **1** 195–200
[6] Spekkens R W 2005 Contextuality for preparations, transformations, and unsharp measurements *Phys. Rev.* A **71** 052108
[7] Cabello A S 2012 Specker's fundamental principle of quantum mechanics arXiv:1212.1756
[8] Liang Y-C, Spekkens R W and Wiseman H M 2011 Specker's parable of the overprotective seer: a road to contextuality, nonlocality and complementarity *Phys. Rep.* **506** 1–39
[9] Specker E P 1960 Die logik nicht gleichzeitig entscheidbarer aussagen *Dialectica* **14** 239–46
Reprinted in Specker E P 1990 *Selecta* (Basel: Birkhäuser) pp 175–82
Stairs A 1975 The logic of propositions which are not simultaneously decidable *The Logico-Algebraic Approach to Quantum Mechanics Volume I: Historical Evolution* ed C A Hooker (Dordrecht: Reidel) pp 135–40 (First Engl. Transl.)
Seevinck M P *The Logic of Non-Simultaneously Decidable Propositions* (Second Engl. Transl.)arXiv:1103.4537
[10] Foulis D and Randall C 1970 An approach to empirical logic *Am. Math. Mon.* **77** 363–74
[11] Foulis D and Randall C 1974 The empirical logic approach to the physical sciences *Foundations of Quantum Mechanics and Ordered Linear Spaces* ed A Hartkämper and H von Neumann (Berlin: Springer)
[12] Foulis D and Randall C 1981 What are quantum logics and what ought they to be? *Current Issues in Quantum Logic* ed E G Beltrametti and B C van Fraassen (NY: Plenum)
[13] Foulis D and Randall C 1981 Empirical logic and tensor products *Interpretations and Foundations of Quantum Theory* ed H Neumann (Mannheim: B. I. Wisssenschaft)
[14] Gunson J 1967 On the algebraic structure of quantum mechanics *Commun. Math. Phys.* **6** 262–85
[15] Mielnik B 1968 Geometry of quantum states *Commun. Math. Phys.* **9** 55–80
[16] Ludwig G 1964 Versuch einer axiomatischen Grundlegung der Quantenmechanik und allgemeinerer physikalischer theorien *Z. Phys.* **181** 233–60
Ludwig G 1967 Attempt of an axiomatic foundation of quantum mechanics and more general theories II *Commun. Math. Phys.* **4** 331–48
Ludwig G 1968 Attempt of an axiomatic foundation of quantum mechanics and more general theories III *Commun. Math. Phys.* **9** 1–12
[17] Piron C 1976 *Foundations of Quantum Physics* (Reading, Massachusetts: Benjamin)
[18] Mackey G W 1963 *Mathematical Foundations of Quantum Mechanics* (New York: Benjamin)
Mackey G W 2004 *Mathematical Foundations of Quantum Mechanics* (New York: Dover)
[19] Wilce A 2012 Four and a half axioms for finite-dimensional quantum probability *Probability in Physics* ed Y Ben-Menahem and M Hemmo (Berlin: Springer) pp 281–98
[20] Ludwig G 1985 *An Axiomatic Basis for Quantum Mechanics I: derivation of Hilbert Space Structure* (Berlin: Springer)
[21] Barnum H 2002 Quantum information processing and quantum logic: towards mutual illumination *2001 Meeting of the IQSA (Cesena/Cesenatico, Italy)* arXiv:quant-ph/0205129
[22] Barnum H 2002 Quantum information processing, operational quantum logic, convexity, and the foundations of physics *Stud. Hist. Phil. Mod. Phys.* **34** 449–58
[23] Cassinelli G and Lahti P 2016 An axiomatic basis for quantum mechanics *Found. Phys.* **46** 1–33
[24] Alfsen E M and Shultz F W 2003 *Geometry of State Spaces of Operator Algebras* (Boston: Birkhäuser)
[25] Alfsen E M and Shultz F W 1980 State spaces of Jordan algebras *Acta Math.* **144** 267–305
[26] Varadarajan V S 1985 *Geometry of Quantum Theory* 2nd edn (New York: Springer)

[17] Here 'y covers x' means '$y \geqslant x$, $y \neq x$, and there is no z distinct from x and y such that $y \geqslant z \geqslant x$, i.e. '$y$ is above x with nothing in between. An atom is an element that covers 0.

[18] The results in [20] were mostly obtained in a series of papers in the late 1960s and early 1970s.

[19] Both Piron and Ludwig also made an atomicity assumption (which may be considered more technical than substantive, and always holds in finite dimension) and also assumed lattice dimension 4 or greater, so Hilbert spaces of dimension 3 or less were not dealt with, nor were spin factors or the exceptional Jordan algebra. These low-dimensional cases also satisfy Piron's, and Ludwig's premises, but a theorem ruling out other instances satisfying them appears to be lacking.

[27] Wilce A 2002 Quantum logic and probability theory *The Stanford Encyclopedia of Philosophy (Spring 2017 Edition)* Edward N. Zalta (ed.) http://plato.stanford/edu/entries/qt-quantlog
[28] Horodecki M and Oppenheim J 2013 Fundamental limitations for quantum and nanoscale thermodynamics *Nat. Comm.* **4** 2059
[29] Gour G, Müller M P, Narasimhachar V, Spekkens R W and Halpern N Y 2015 The resource theory of informational nonequilibrium in thermodynamics *Phys. Rep.* **583** 1–58
[30] Brandao F G S Ł, Horodecki M, Ng N H Y, Oppenheim J and Wehner S 2015 The second laws of quantum thermodynamics *Proc. Natl. Acad. Sci.* **112** 3275
[31] Barnum H, Barrett J, Krumm M and Müller M P 2015 Entropy, majorization and thermodynamics in general probabilistic theories *Proc. 12th Int. Workshop on Quantum Physics and Logic* EPTCS 195, pp 43–58 arXiv:1508.03107
[32] Schilpp P A 1979 *Albert Einstein: Autobiographical Notes* Centennial edn Open Court Publishing) p 31
As quoted by Howard D and Stachel J 2000 *Einstein: The Formative Years, 1879–1909* (*Einstein Studies* vol 8) (Boston: Birkhäuser)
[33] Eddington A S 1915 *The Nature of the Physical World* ch 4 (Cambridge: Cambridge University Press)
[34] Krumm M 2015 Thermodynamics and the structure of quantum theory as a generalized probabilistic theory *Master Thesis* Heidelberg University arXiv:1508.03299
[35] Müller M P, Dahlsten O C O and Vedral V 2012 Unifying typical entanglement and coin tossing: on randomization in probabilistic theories *Commun. Math. Phys.* **316** 441–87
[36] Müller M P, Oppenheim J and Dahlsten O C O 2012 The black hole information problem beyond quantum theory *J. High. Energy Phys.* JHEP09(2012)116
[37] Popescu S, Short A J and Winter A 2006 Entanglement and the foundations of statistical mechanics *Nat. Phys.* **2** 754–8
[38] Goldstein S, Lebowitz J L, Tumulka R and Zanghi N 2006 Canonical typicality *Phys. Rev. Lett.* **96** 050403
[39] Müller M P, Adlam E, Masanes L and Wiebe N 2015 Thermalization and canonical typicality in translation-invariant quantum lattice systems *Commun. Math. Phys.* **340** 499–561
[40] Page D N 1993 Information in black hole radiation *Phys. Rev. Lett.* **71** 3743–6
[41] von Neumann J 1932 *Mathematische Grundlagen der Quantenmechanik* (Berlin: Springer)
von Neumann J 1955 *Mathematical Foundations of Quantum Mechanics* (Princeton: Princeton University Press) (Engl. Transl.)
[42] Petz D 2001 Entropy, von Neumann and the von Neumann entropy *John von Neumann and the Foundations of Quantum Physics* ed M Rédei and M Stöltzner (Dordrecht: Kluwer) arXiv:math-ph/0102013v1
[43] Barnum H, Müller M P and Ududec C 2014 Higher-order interference and single-system postulates characterizing quantum theory *New J. Phys.* **16** 123029
[44] Carnot S 1897 *Reflections on the Motive Power of Heat and on Machines Fitted to Develop Power* (London: Chapman and Hall) (Engl. Transl. by Sir W Thompson (Lord Kelvin))
[45] Chiribella G and Scandolo C M 2015 Operational axioms for diagonalizing states *Proc. 12th Int. Workshop on Quantum Physics and Logic* EPTCS 195, pp 96–115 arXiv:1506.00380
[46] Hänggi E and Wehner S 2013 A violation of the uncertainty principle implies a violation of the second law of thermodynamics *Nat. Comm.* **4** 1670
[47] Chiribella G and Scandolo C M 2015 Entanglement and thermodynamics in general probabilistic theories *New J. Phys.* **17** 103027
[48] Scandolo C M 2014 Entanglement and thermodynamics in general probabilistic theories *Master's Thesis* Università degli Studi di Padova, Italy http://tesi.cab.unipd.it/46015/1/Scandolo_carlo_maria.pdf
[49] Barnum H, Graydon M A and Wilce A 2015 Some nearly quantum theories *Proc. 12th Int. Workshop on Quantum Physics and Logic* EPTCS 195, pp 59–70 arXiv:1507.06278
[50] Nielsen M and Chuang I 2010 *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press) 10th Anniversary edition published 2010, 6th printing 2014
[51] Hardy L 2001 Quantum theory from five reasonable axioms arXiv:quant-ph/0101012
[52] Barrett J 2007 Information processing in generalized probabilistic theories *Phys. Rev.* A **75** 032304
[53] Fuchs C A 2002 Quantum mechanics as quantum information (and only a little more) *Quantum Theory: Reconstruction of Foundations* ed A Khrenikov (Växjo: Växjo University Press)
[54] Dakić B and Brukner C 2011 Quantum theory and beyond: Is entanglement special? *Deep Beauty: Understanding the Quantum World through Mathematical Innovation* ed H Halvorson (Cambridge: Cambridge University Press) arXiv:0911.0695
[55] Masanes L and Müller M P 2011 A derivation of quantum theory from physical requirements *New J. Phys.* **13** 063001
[56] Chiribella G, D'Ariano G M and Perinotti P 2010 Probabilistic theories with purification *Phys. Rev.* A **81** 062348
[57] Chiribella G, D'Ariano G M and Perinotti P 2011 Informational derivation of quantum theory *Phys. Rev.* A **84** 012311
[58] Hardy L 2011 Reformulating and reconstructing quantum theory arXiv:1104.2066
[59] Masanes L, Müller M P, Augusiak R and Pérez-García D 2013 Existence of an information unit as a postulate of quantum theory *Proc. Natl. Acad. Sci. USA* **110** 16373
[60] Höhn P A 2014 Toolbox for reconstructing quantum theory from rules on information acquisition arXiv:1412.8323
[61] Höhn P A and Wever C S P 2017 Quantum theory from questions *Phys. Rev.* A **95** 012102
[62] Ududec C 2012 Perspectives on the formalism of quantum theory *PhD Thesis* University of Waterloo
[63] Pfister C 2011 One simple postulate implies that every polytopic state space is classical *Master Thesis* ETH, Zürich
[64] Müller M P and Ududec C 2012 The structure of reversible computation determines the self-duality of quantum theory *Phys. Rev. Lett.* **108** 130401
[65] Rényi A 1961 On measures of entropy and information *Proc. 4th Berkeley Symp. on Mathematical Statistics and Probability 1: Contributions to the Theory of Statistics* (Berkeley: University of California Press) pp 547–61
[66] Short A J and Wehner S 2010 Entropy in general physical theories *New J. Phys.* **12** 033023
[67] Barnum H, Barrett J, Clark L O, Leifer M, Spekkens R, Stepanik N, Wilce A and Wilke R 2010 Entropy and information causality in general probabilistic theories *New J. Phys.* **12** 033024
[68] Kimura G, Nuida K and Imai H 2010 Distinguishability measures and entropies for general probabilistic theories *Rep. Math. Phys.* **66** 175
[69] Hein C A 1979 Entropy in operational statistics and quantum logic *Found. Phys.* **9** 751–86
[70] Popescu S 2014 Nonlocality beyond quantum mechanics *Nat. Phys.* **10** 264–70
[71] van Dam W 2013 Implausible consequences of superstrong nonlocality *Nat. Comput.* 1–4
[72] Navascués M and Wunderlich H 2010 A glance beyond the quantum model *Proc. R. Soc.* A **466** 881–90
[73] Pawlowski M, Paterek T, Kaszlikowski D, Scarani V, Winter A and Zukowski M 2009 Information causality as a physical principle *Nature* **461** 1101–4

[74] Dahlsten O C O, Lercher D and Renner R 2012 Tsirelson's bound from a generalized data processing inequality *New J. Phys.* **14** 063024
[75] Navascués M, Guryanova Y, Hoban M J and Acín A 2015 Almost quantum correlations *Nat. Comm.* **6** 6288
[76] Sorkin R D 1994 Quantum mechanics as quantum measure theory *Mod. Phys. Lett.* A **9** 3119–27
[77] Craig D, Dowker F, Henson J, Major S, Rideout D and Sorkin R D 2007 A bell inequality analog in quantum measure theory *J. Phys.* A **40** 501–23
[78] Jacobson T 1995 Thermodynamics of spacetime: the Einstein equation of state *Phys. Rev. Lett.* **75** 1260–3
[79] Ryu S and Takayanagi T 2006 Holographic derivation of entanglement entropy from AdS/CFT *Phys. Rev. Lett.* **96** 181602
[80] Horowitz G T and Maldacena J 2004 The black hole final state *J. High. Energy Phys.* JHEP02(2004)008
[81] Barnum H, Graydon M A and Wilce A 2016 *Some Nearly Quantum Theories* ed P. Selinger and J. Vicary Proceedings of the 12th International Workshop on Quantum Physics and Logic, EPTCS 195, pp. 5970 (arXiv:1507.06278)
[82] Brunner N, Kaplan M, Leverrier A and Skrzypczyk P 2014 Dimension of physical systems, information processing, and thermodynamics *New J. Phys.* **16** 123050
[83] Popescu S and Rohrlich D 1994 Quantum nonlocality as an axiom *Found. Phys.* **24** 379–85
[84] Davies E B and Lewis J T 1970 An operational approach to quantum probability *Commun. Math. Phys.* **17** 239–60
[85] Barnum H, Barrett J, Leifer M and Wilce A 2007 A generalized no-broadcasting theorem *Phys. Rev. Lett.* **99** 240501
[86] Janotta P and Hinrichsen H 2014 Generalized probability theories: What determines the structure of quantum physics? *J. Phys. A: Math. Theor.* **47** 323001
[87] Pfister C and Wehner S 2013 An information-theoretic principle implies that any discrete physical theory is classical *Nat. Comm.* **4** 1851
[88] Mielnik B 1974 Generalized quantum mechanics *Commun. Math. Phys.* **37** 221–56
[89] Webster R 1994 *Convexity* (Oxford: Oxford University Press)
[90] Schwabl F 2006 *Statistische Mechanik* (Berlin: Springer)
[91] Acín A, Fritz T, Leverrier A and Sainz A B 2015 A combinatorial approach to nonlocality and contextuality *Commun. Math. Phys.* **334** 533–628
[92] Bennett C H 1982 The thermodynamics of computation–a review *Int. J. Theor. Phys.* **21** 905–40
[93] Ududec C, Barnum H and Emerson J 2011 Three slit experiments and the structure of quantum theory *Found. Phys.* **41** 396–405
[94] Piron C 1964 Axiomatique quantique *Helvetica Physica Acta* Helvetica Physica Acta **37** 439–68
[95] Henson J 2015 Bounding quantum contextuality with lack of third-order interference *Phys. Rev. Lett.* **114** 220403
[96] Lee C M and Selby J H 2017 Higher-order interference in extensions of quantum theory *Foundations of Physics* **47** 89–112 Higher-order interference in extensions of quantum theory arXiv:1510.03860
[97] Sinha U, Couteau C, Jennewein T, Laflamme R and Weihs G 2010 Ruling out multi-order interference in quantum mechanics *Science* **329** 418
[98] Sinha U, Couteau C, Medendorp Z, Söllner I, Laflamme R, Sorkin R and Weihs G 2009 Testing borns rule in quantum mechanics with a triple slit experiment *AIP Conf. Proc.* **1101** 200
[99] Söllner I, Gschösser B, Mai P, Pressl B, Vörös Z and Weihs G 2012 Testing borns rule in quantum mechanics for three mutually exclusive events *Found. Phys.* **42** 742–51
[100] Kauten T, Keil R, Kaufmann T, Pressl B and Weihs G 2017 Obtaining tight bounds on higher-order interferences with a 5-path interferometer *New J. Phys.* **9** 033017
[101] Hickmann J M, Fonseca E J S and Jesus-Silva A J 2011 Born's rule and the interference of photons with orbital angular momentum by a triangular slit *EPL* **96** 64006
[102] Park D K, Moussa O and Laflamme R 2012 Three path interference using nuclear magnetic resonance: a test of the consistency of born's rule *New J. Phys.* **14** 113025