


## Contextuality of General Probabilistic Theories

Farid Shahandeh<sup>1</sup>\*

*Department of Physics, Swansea University, Singleton Park, Swansea SA2 8PP, United Kingdom*

 (Received 27 January 2020; revised 24 April 2020; accepted 11 January 2021; published 25 February 2021)

Generalized contextuality refers to our inability of explaining measurement statistics using a context-independent probabilistic and ontological model. On the other hand, measurement statistics can also be modeled using the framework of general probabilistic theories (GPTs). Here, starting from a construction of GPTs based on a Gleason-type theorem, we fully characterize these structures with respect to their permission and rejection of generalized (non)contextual ontological models. It follows that in any GPT construction the three insistence of (i) the no-restriction hypothesis, (ii) the ontological noncontextuality, and (iii) multiple nonrefinable measurements for any fixed number of outcomes are incompatible. Hence, any GPT satisfying the no-restriction hypothesis is ontologically noncontextual if and only if it is simplicial. We give a detailed discussion of GPTs for which the no-restriction hypothesis is violated, and show that they can always be considered as subtheories (subGPTs) of GPTs satisfying the hypothesis. It is shown that subGPTs are ontologically noncontextual if and only if they are subtheories of simplicial GPTs of the same dimensionality. Finally, we establish as a corollary the necessary and sufficient condition for a single resourceful measurement or state to promote an ontologically noncontextual (i.e., classical) general probabilistic theory to an ontologically contextual (i.e., nonclassical) one under the no-restriction hypothesis.

DOI: [10.1103/PRXQuantum.2.010330](https://doi.org/10.1103/PRXQuantum.2.010330)

### I. INTRODUCTION

A promise of classical theories is to give a probabilistic account for the statistics resulted from prepare-and-measure experiments on single physical systems without relying upon their particular operational procedure, or *context*. Contextuality [1–4], and more recently, generalized contextuality [5], refer to no-go theorems dismissing context-independent classical models for measured statistics. Hence, a plausible definition of nonclassicality of statistics is through their generalized contextuality [6,7].

Besides probabilistic (or *ontological*) models, statistics from operational procedures can also be modeled using the framework of *general probabilistic theories* (GPTs), an example of which is quantum theory [8–14]. Nonclassical data thus mean that none of their potential GPT explanations can be translated into a noncontextual ontological model [5]. Recently, Kunjwal and Spekkens [15] and Schmid *et al.* [16] provided noncontextuality inequalities for detection of possible nonclassical statistics. In this

paper, we determine which GPTs for physical phenomena allow and which ones disallow noncontextual ontological models. More specifically, we prove a Gleason-type theorem for construction of GPTs and show that, in finite dimensions, any GPT that satisfies the no-restriction hypothesis [10] must possess simplex sets of nonrefinable effects and states to be ontologically noncontextual. We then discuss the scenario in which the no-restriction hypothesis is removed.

The practical significance of our analysis lies within a markedly interesting context. Referring to the scenario of quantum computations with Clifford circuits, stabilizer input states, and Pauli measurements, it is well known that such computations authorize a classical model making them classically efficiently simulatable [17–20]. This possibility is removed by providing only a *single* suitable nonstabilizer input state (or measurement) that “magically” enables fault-tolerant universal quantum computing [20–24]. Thus, there are scenarios in which only a single extra preparation or measurement procedure is *resourceful* in that it simultaneously gives rise to two phenomena. First, it generates data that render a classical model impossible. Second, it causes a significant improvement in the performance of some information-processing protocols. Our analysis here is motivated by the first phenomenon, which is a prerequisite of the second.

The study of resources for information-processing purposes commonly begins with assuming an underlying

\*shahandeh.f@gmail.com

*Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.*

theory. Thinking of quantum theory, this is beautifully done within the formalism of quantum resource theories [25], such as entanglement [26], athermality [27–32], coherence [33–36], asymmetry [37–39], non-Markovianity [40], and dynamical correlations [41]. Here, we instead relax the assumption of a specific underlying theory by adopting the generic formalism of GPTs and use our first result to obtain the necessary and sufficient criterion to distinguish between classicality and non-classicality of a single resourceful measurement or state leading to potential nonclassical information-processing advantages [7].

## II. ONTOLOGICAL MODELS

Throughout this paper, we are interested only in the prepare-and-measure experiments on single systems that can be described in finite dimensions. Operationally, the primitive elements of our physical description in any such experiment are the *laboratory prescriptions* for preparations and measurements forming the collections  $\mathcal{P}:=\{\mathbf{P}_k\}$  and  $\mathcal{M}:=\{\mathbf{M}_j\}$ , respectively [5]. The fundamental goal in theoretical physics is to establish assignments between these elements and mathematical objects endowed with a set of rules to determine the outcome probabilities in each measurement. A first generic step is to meaningfully assign (not necessarily scalar) “sizes,” called *measures*, to measurement outcomes. Given the finite set  $\Omega$  of all outcomes, a set  $\omega$  of its subsets on which such assignments are well defined is called the  $\sigma$  algebra of events and the pair  $(\Omega, \omega)$  is named a *measurable space*; see Appendix A for a brief review.

Ontological models go beyond the minimal promise of theoretical physics by assuming an underlying *ontic* variable space  $\Upsilon$  and assigning physical phenomena to *elements of reality*. In this process, the operational elements, i.e., preparations and measurements, correspond most generally to probabilistic preparations and measurements of the ontic variable and thus, they are represented by probability distributions and indicator functions over  $\Upsilon$ , respectively. More precisely, one defines also a  $\sigma$  algebra  $\nu$  on  $\Upsilon$  and designates “sizes” to members of both  $\omega$  and  $\nu$  that are probabilities, also known as *probability measures*. Given the collections  $\mathcal{V}$  and  $\mathcal{Q}$  of all probability measures on  $(\Upsilon, \nu)$  and  $(\Omega, \omega)$ , respectively, the ontological model then hypothesizes the existence of *convex linear* maps  $\mu: \mathcal{P} \rightarrow \mathcal{V}$  and  $\xi: \mathcal{M} \rightarrow \mathcal{Q}$  that assign the *ontic state*  $\mu_{\mathbf{P}}$  to the preparation procedure  $\mathbf{P}$  and the *ontic measurement*  $\xi_{\mathbf{M}}$  to the measurement procedure  $\mathbf{M}$  [5]. Thus, for each preparation,

$$\mu_{\mathbf{P}} : \nu \rightarrow [0, 1] \quad \text{and} \quad \int_{\Upsilon} d\lambda \mu_{\mathbf{P}}(\lambda) = 1, \quad (1)$$

and for each measurement,

$$\xi_{\mathbf{M}} : \omega \times \Upsilon \rightarrow [0, 1] \quad \text{and} \quad \xi_{\mathbf{M}}(\Omega|\lambda) = 1 \quad \forall \lambda \in \Upsilon. \quad (2)$$

Then, the probability of a particular event  $X$  in a measurement  $\mathbf{M}$  given the preparation  $\mathbf{P}$  can be obtained via Bayes’ rule,

$$p(X|\mathbf{P}, \mathbf{M}) = \int_{\Upsilon} d\lambda \mu_{\mathbf{P}}(\lambda) \xi_{\mathbf{M}}(X|\lambda). \quad (3)$$

## III. GENERAL PROBABILISTIC THEORIES

A second approach to the abstraction of operational scenarios is known as GPTs. Their constructions begin with assuming a vector space  $\mathcal{V}$  whose elements can (at least partially) be ordered and on which an inner product  $\langle \cdot, \cdot \rangle$  can be defined [8–14]. In the example of quantum theory, this vector space is the Banach space  $\mathcal{L}(\mathcal{H})$  of all bounded linear Hermitian operators on a Hilbert space  $\mathcal{H}$  with the usual Hilbert-Schmidt inner product. In contrast to ontological models, in this case, to each measurement event  $X \in \omega$  we assign a *vector* as its “size”; Appendix A. Hence, instead of a probability measure we have a *probability vector-valued measure* (PVVM) which is a function  $E: \omega \rightarrow \mathcal{V}$  satisfying (i)  $E(X) \geq 0$  for all  $X \in \omega$ , (ii)  $E(\Omega) = U$  for a fixed nonzero element  $U \in \mathcal{V}$  called the *unit* element, and (iii)  $E(\cup_i X_i) = \sum_i E(X_i)$  for all sequences of disjoint events  $X_i \in \omega$ . An easy way to make sense of these conditions is by comparing a PVVM to a probability measure where after the substitutions  $\mathcal{V} \mapsto \mathbb{R}$  and  $U \mapsto 1$  the former simply reduces to the latter; see Appendix A. Each vector  $E(X_i)$  is called an *effect* where their collection is denoted by  $\mathcal{E}$ . The familiar quantum counterpart of a PVVM is a *positive operator-valued measure* (POVM) [42,43].

Suppose that along with a measurement having  $E(X)$  as an effect the experimenter tosses a (biased) coin with probability  $p$  for heads and accepts the occurrences of  $X$  only if the coin is heads. Assuming that the product rule of probability holds, the effect corresponding to the accepted events is given by  $pE(X)$ . Hence, it is reasonable to assume that given any operationally legitimate effect  $E(X)$ , the effect  $pE(X)$  for any real number  $p \in [0, 1]$  is also allowed. Further, suppose that  $E_1$  and  $E_2$  are two PVVMs and our experimenter performs  $E_1$  if the coin is heads and  $E_2$  otherwise. The resulting PVVM is thus  $E_p := pE_1 + (1-p)E_2$  implying that the set of all effects  $\mathcal{E}$  is convex. In order for the latter to be a well-defined summation,  $E_1$  and  $E_2$  have to possess a common domain. Consequently, throughout this paper we assume that the event space  $\omega$  is fixed. Finally, we assume that  $\mathcal{E}$  spans  $\mathcal{V}$ . These allow us to state a Gleason-type theorem for GPTs as follows.

**Theorem 1.** *Any generalized probability measure  $q: \mathcal{E} \rightarrow [0, 1]$  satisfying (i)  $q[E(X)] \geq 0$  for all effects  $E(X) \in \mathcal{E}$ , (ii)  $q(U) = 1$ , and (iii)  $q[\sum_i E(X_i)] = \sum_i q[E(X_i)]$  for all*

sequences of effects in  $\mathcal{E}$  that satisfy  $\sum_i E(X_i) \leq U$ , must be of the form  $q(A) = \langle A, B \rangle$  for all  $A \in \mathcal{V}$ , for a unique  $B \in \mathcal{V}$ , which is normalized in the sense that  $\langle U, B \rangle = 1$ .

As presented in Appendix B, Theorem 1 is simply proven by extending probability measures on  $\mathcal{E}$  to the whole  $\mathcal{V}$  and using Riesz's representation theorem.

Given effect space  $\mathcal{E}$ , the state space of the GPT can be delineated using Theorem 1 in conjunction with the *no-restriction hypothesis* [10], i.e., that given PVVM space  $\mathcal{E}$  all definable probability measures ( $q$ 's) on it correspond to physically valid states, as

$$\mathcal{S} := \{ \varrho \in \mathcal{V} \mid \langle E(X), \varrho \rangle \geq 0 \forall E(X) \in \mathcal{E}, \langle U, \varrho \rangle = 1 \}. \quad (4)$$

We denote a GPT by its pair of PVVM and state space as  $T := (\mathcal{E}, \mathcal{S})$ .

Here, instead of *a priori* assuming the probability rule of GPTs, we derive it from a set of reasonable assumptions following in the footsteps of Gleason [44], Busch [45], and Caves *et al.* [46]. Moreover, it is argued, for example in Refs. [11, 47, 48], that the no-restriction hypothesis is of no physical basis and thus, it is desirable to drop it from GPT constructions. We note that, such a ‘‘relaxation’’ comes at a price: one has to *assume a priori* also the state assignments, meaning trading one assumption for another. Indeed, there is no objection to not imposing the no-restriction hypothesis, however, we show in Appendix C that any GPT that does not satisfy the no-restriction hypothesis is obtained as a *subtheory* of possibly (infinitely) many GPTs that do satisfy it by imposing appropriate further constraints, thus named a subGPT. To give an example, consider quantum mechanics as a specific GPT that satisfies the no-restriction hypothesis and from which the subGPT of Gaussian quantum mechanics is obtained by restricting the effect and state spaces to those elements possessing Gaussian Wigner representations [49]. Notably, quantum theory is *not* the unique GPT containing Gaussian quantum mechanics, for it is also a subtheory of classical statistical (or Liouville) mechanics [49].

Within the rest of this paper, we use GPT and subGPT to distinguish between theories that do and do not comply with the no-restriction hypothesis, respectively.

#### IV. BROAD (NON)CONTEXTUALITY

Irrespective of which approach is adopted to explain experimental data, be it an ontological model or a (sub)GPT, *broad noncontextuality* (as defined below) is a desirable hypothesis about the description. Beginning with the *statistical equivalence* assumption [5, 42], two preparations  $P_1, P_2 \in \mathcal{P}$  are statistically indiscernible and equivalent,  $P_1 \cong P_2$ , if and only if for every measurement procedure  $M \in \mathcal{M}$  and every event  $X \in \omega$  it holds that  $p(X|P_1, M) = p(X|P_2, M)$ . Similarly, two measurements  $M_1, M_2 \in \mathcal{M}$  are statistically indiscernible and

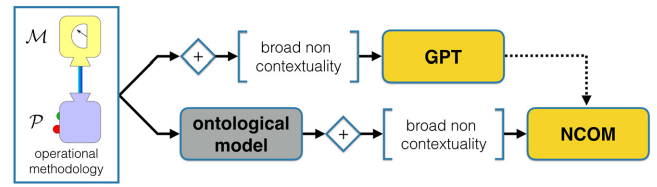


FIG. 1. The diagrammatic representation of different model constructions for operational descriptions of experiments. The dotted arrow represents the detour approach for building NCOMs for GPTs. A GPT that admits a NCOM is called ontologically noncontextual.

equivalent,  $M_1 \cong M_2$ , if and only if for every preparation procedure  $P \in \mathcal{P}$  and every event  $X \in \omega$  it holds that  $p(X|P, M_1) = p(X|P, M_2)$ . These relations partition the collections of preparations and measurements into equivalence classes  $e(P)$  and  $e(M)$  for each preparation  $P$  and measurement  $M$ . The particular way in which a state or measurement is experimentally realized corresponds to an element within an equivalence class and it is called a *context*. The *broad noncontextuality hypothesis* states that our models of physical phenomena aiming *only* at reproducing the statistics should depend only on equivalence classes rather than individual contexts, because statistics do not carry any information about the latter; see, e.g., Ref. [46].

An ontological model, which is noncontextual in the broad sense is called a *noncontextual ontological model (NCOM)* and satisfies

$$\begin{aligned} P_1 \cong P_2 &\Leftrightarrow \mu_{P_1} = \mu_{P_2}, \\ M_1 \cong M_2 &\Leftrightarrow \{\xi_{M_1}(X|\lambda)\} = \{\xi_{M_2}(X|\lambda)\}. \end{aligned} \quad (5)$$

Similarly, broad noncontextuality of (sub)GPTs reads as

$$\begin{aligned} P_1 \cong P_2 &\Leftrightarrow P_1, P_2 \mapsto \varrho, \\ M_1 \cong M_2 &\Leftrightarrow M_1, M_2 \mapsto \{E(X)\}. \end{aligned} \quad (6)$$

Note that, broad noncontextuality is built in to our GPT and subGPT constructions. In general, however, it is possible to construct (sub)GPTs that do not respect the broad noncontextuality hypothesis by allowing for context-dependent effect and state assignments. Importantly, it follows from our constructions here that every NCOM is a (sub)GPT while the converse is not true. In this paper, our aim is to determine which (sub)GPTs do and which ones do not admit an NCOM; see Fig. 1.

#### V. ONTOLOGICAL (NON)CONTEXTUALITY OF (SUB)GPTS

As we infer from experiments in the quantum regime, that to take a *direct* route and build a NCOM to describe all possible physical experiments seems very unlikely. Yet we may ask if it is possible to take a detour, as shown

in Fig. 1, and construct ontological models of (sub)GPTs, noting that, with appropriate care, such models will inherit the noncontextuality from the theory leading to NCOMs. For quantum theory as a GPT the answer to our question is in negative [6,50]. In the case of a generic (sub)GPT, we replace the preparation and measurement procedures  $\mathbf{P}$  and  $\mathbf{M}$  in Eqs. (1)–(3), with their representatives in the theory,  $\varrho \in \mathcal{S}$  and  $\{E(X)\} \subset \mathcal{E}$ , respectively. Hence, there should exist *injective* maps  $\eta: \mathcal{S} \rightarrow \mathcal{Y}$  and  $\zeta: \mathcal{E} \rightarrow \mathcal{Q}$  that assign the *unique* ontic state  $\eta_\varrho$  and ontic measurement  $\zeta_E$  to each state vector  $\varrho$  and PVVM  $E$ , respectively, such that for all  $\lambda \in \Upsilon$  and all events  $X \in \omega$ ,

$$\eta_\varrho(\lambda) \geq 0, \quad \zeta_E(X|\lambda) \in [0,1], \quad (7)$$

and satisfy

$$\int_{\Upsilon} d\lambda \eta_\varrho(\lambda) = 1, \quad \text{and} \quad \forall \lambda \quad \zeta_E(\Omega|\lambda) = 1. \quad (8)$$

The probability of a particular event  $X$  in a measurement  $\mathbf{M}$  given the preparation  $\mathbf{P}$  should then be obtained as

$$p(X|\mathbf{P},\mathbf{M}) = p(X|\varrho, E) = \int_{\Upsilon} d\lambda \eta_\varrho(\lambda) \zeta_E(X|\lambda). \quad (9)$$

Within recent literature a *noncontextual (sub)theory* usually refers to a (sub)theory that admits an NCOM [5]. For clarity, here we call such a (sub)theory *ontologically noncontextual* and reserve the term *noncontextuality* for the broader notion. We note also that any ontologically noncontextual (sub)theory is noncontextual while the converse is not true. For instance, quantum theory satisfies broad noncontextuality hypothesis but, it is not ontologically noncontextual as it contains effects and states that do not possess a unique convex decomposition in terms of the respective extremal elements [6,50]. The latter is clarified by the following analysis.

### A. GPTs

We now analyze ontological noncontextuality of GPTs. Recall that the maps  $\mu$  and  $\xi$ , and hence  $\eta$  and  $\zeta$ , are convex linear. Using the fact that  $\mathcal{S}$  and  $\mathcal{E}$  both span  $\mathcal{V}$ ,  $\eta$  and  $\zeta$  can uniquely be extended to the whole space  $\mathcal{V}$ . Then, using Riesz’s theorem we find that they must be of the forms  $\eta_\varrho(\lambda) = \langle \varrho, F(\lambda) \rangle$  and  $\zeta_E(X|\lambda) = \langle E(X), D(\lambda) \rangle$  for  $F(\lambda), D(\lambda) \in \mathcal{V}$ . Satisfying Eq. (8) then requires that  $\int_{\Upsilon} d\lambda F(\lambda) = U$  and  $\langle U, D(\lambda) \rangle = 1$  for all  $\lambda \in \Upsilon$ . Thus,  $\mathcal{F} := \{F(\lambda)\}$  resembles a PVVM whereas  $\mathcal{D} := \{D(\lambda)\}$  is a subset of GPT’s state space. The probability rule of Eq. (9) then implies that  $\mathcal{F}$  and  $\mathcal{D}$  are dual frames [50,51] for

representation of vectors in  $\mathcal{V}$  so that

$$\varrho = \int_{\Upsilon} d\lambda \eta_\varrho(\lambda) D(\lambda), \quad \text{and} \quad E(X) = \int_{\Upsilon} d\lambda F(\lambda) \zeta_E(X|\lambda). \quad (10)$$

After taking into account that state vectors are dual to measurement vectors and a few more simple steps (see Appendix E) we find that  $\mathcal{F}$  and  $\mathcal{D}$  must be generating sets of closed convex sets  $\mathcal{E}$  and  $\mathcal{S}$ , respectively, i.e.,  $\mathcal{E} = \text{conv} \mathcal{F}$  and  $\mathcal{S} = \text{conv} \mathcal{D}$ . In light of Eq. (10), we conclude from the latter that the extremal or *pure* states  $\varrho \in \mathcal{D}$  and the extremal or *sharp* effects  $E(X) \in \mathcal{F}$ , i.e., states and effects that cannot convexly be decomposed into other states and effects, must be represented by Dirac delta measures over the ontic space  $\Upsilon$ , that is, (i)  $\mathcal{D} \ni \varrho \mapsto \eta \delta_{\lambda_\varrho}(\lambda)$  for some  $\lambda_\varrho \in \Upsilon$  and (ii)  $\mathcal{F} \ni E(X) \mapsto \zeta \delta_{\lambda_{E(X)}}(\lambda)$  for some  $\lambda_{E(X)} \in \Upsilon$ , where  $\delta_a(\beta)$  equals 0 if  $a \notin \beta$  and equals 1 if  $a \in \beta$  for any measurable subset  $\beta$ .

The condition (i) is referred to as *ontic determinism*, which means a *pure* preparation [10] represented by a pure state determines the ontic value  $\lambda$  completely. Accordingly, we call condition (ii) *outcome determinability*, that is, in *sharp* measurements [52] represented by sharp effects of the theory, specifying the ontic variable determines the outcome of the measurement with certainty. Ontic determinism and outcome determinability enforce that in any ontologically noncontextual GPT state and effect vectors possess *unique* decompositions into nonrefinable extremal elements. Specifically, it is clear from Eq. (10) that requiring a unique ontic representation  $\eta_\varrho(\lambda)$  for a state vector  $\varrho$  simply means a unique convex decomposition into pure states. Considering sharp effects that cannot be refined (or *atomic* effects [10]) and denoting their collection by  $\mathcal{E}_{\text{NR}} \subset \mathcal{F}$ , it follows also that the convex decomposition of any effect into elements of  $\mathcal{E}_{\text{NR}}$  must be unique. The following theorem characterizes GPTs for which these criteria are met and thus, they are ontologically noncontextual. For a detailed proof, see Appendix E.

**Theorem 2.** *A GPT is ontologically noncontextual if and only if its pure states and nonrefinable sharp effects each form a complete basis for the space  $\mathcal{V}$ . Equivalently, the GPT must be simplicial meaning that  $\mathcal{S}$  and  $\text{conv} \mathcal{E}_{\text{NR}}$  are simplexes.*

### B. SubGPTs

As we discuss earlier, subGPT are GPTs with extra restrictions on their effect and state spaces. We now determine the condition for a subGPT to admit a NCOM.

**Theorem 3.** *Any subGPT  $\mathcal{T}_{\text{sub}} = (\mathcal{E}_{\text{sub}}, \mathcal{S}_{\text{sub}})$  over  $\mathcal{V}_{\text{sub}}$  admits a NCOM if and only if it can be thought of as a subtheory of an ontologically noncontextual GPT  $\mathcal{T} = (\mathcal{E}, \mathcal{S})$  over  $\mathcal{V}$  and  $\dim \mathcal{V}_{\text{sub}} = \dim \mathcal{V} = \text{card } \Upsilon$ .*

The proof of Theorem 3 as detailed in Appendix F, similar to that of Theorem 2, uses the extensions of maps  $\eta$  and  $\zeta$  to the whole  $\mathcal{V}_{\text{sub}}$ , which partially explains the condition  $\dim \mathcal{V}_{\text{sub}} = \dim \mathcal{V}$ . Further clarification is also provided via two examples in Appendix G. On one hand, we show that Spekkens' toy model [53] is a four-dimensional subGPT admitting a NCOM in four dimensions and hence considered to be a classical subGPT; cf. Appendix 1. On the other hand, the stabilizer rebit theory is shown to be a three-dimensional subGPT the smallest ontological model admitted by which is four dimensional and thus, ontologically contextual; cf. Appendix 2.

From our discussion, it follows that the stabilizer rebit subGPT cannot be regarded classical despite the fact that it is classically simulatable in four dimensions. The caveat of the four-dimensional ontological model is that the excess dimension gives room for multiple ontic states that are statistically indiscernible under the subGPT's measurements, hence violating the broad noncontextuality hypothesis of Eq. (5), specifically,  $\mathbf{P}_1 \cong \mathbf{P}_2 \not\Rightarrow \mu_{\mathbf{P}_1} = \mu_{\mathbf{P}_2}$ ; see Appendix 2. Intuitively, this is the case whenever  $\dim \mathcal{V} > \dim \mathcal{V}_{\text{sub}}$ . Classicality of (sub)GPTs thus requires both their classical simulatability and that there are no in-principle inaccessible parameters at the ontological level, i.e., the (sub)GPT is ontologically complete. The latter is in accordance with Leibniz's methodological principle of the ontological identity of empirical indiscernibles [54,55]. We emphasize that this constraint does *not* conflict with NCOM approaches to experimental data presented, e.g., in Refs. [7] and [16]; see Appendix H for a further discussion.

## VI. RESOURCES AND CONTEXTUALITY IN GPTS

We now consider the concept of resources at a fundamental level. Imagine experiments with sets of preparations  $\mathcal{P}$  and measurements  $\mathcal{M}$ , and an experimenter who has devised a GPT  $\mathcal{T} = (\mathcal{E}, \mathcal{S})$  capable of explaining the statistics obtained in the experiments. We call the set of all possible measurements  $\mathcal{E}$  the *free measurements* and the set of all possible preparations  $\mathcal{S}$  the *free states*. We remark that in quantum resource theories [25–41], free states and measurements are subsets of all possible states and measurements. Here, in contrast, all possible preparations and measurements are defined to be free and nonfree (or *resourceful*) ones are those yet to be discovered and thus not specified in the GPT.

Now, suppose that the experimenter discovers another preparation or measurement procedure that was not previously known to exist. Naturally, the experimenter has to come up with a new theory  $\mathcal{T}^* := (\mathcal{E}^*, \mathcal{S}^*)$  to reproduce also the new measurement data. We ask how  $\mathcal{T}^*$  compares to  $\mathcal{T}$  in terms of (non)classicality. We consider only the case wherein the extended theory  $\mathcal{T}^*$  is required to satisfy the no-restriction hypothesis. Then, by Theorem 2, given the

new sets of measurements and preparations,  $\mathcal{M}^*$  and  $\mathcal{P}^*$ , if the nonrefinable sharp measurements of the new GPT form a complete set of PVVMs over a vector space  $\mathcal{V}^*$ , then the GPT will be ontologically noncontextual. In such scenarios, even though the newly discovered element is a resource, all the measured statistics can be explained in classical terms. Hence, we call such a bonus procedure a *classical* resource. On the other hand, a resourceful element may enforce a nonclassical extension of the older theory, where it is called a *nonclassical* resource. The conditions for a single resource to dictate the use of a nonclassical model for an explanation of possible statistics is provided below and proved in Appendix I.

**Theorem 4.** *Suppose that the set of free measurements  $\mathcal{M}$  and preparations  $\mathcal{P}$  are represented by PVVMs  $\mathcal{E}$  and states  $\mathcal{S}$  in some classical GPT  $\mathcal{T} = (\mathcal{E}, \mathcal{S})$ . Given a single nonrefinable bonus measurement  $\mathbf{M}^*$  (preparation  $\mathbf{P}^*$ ), the following are equivalent: (i)  $\mathcal{T}^*$  is ontologically contextual; (ii)  $\mathbf{M}^*$  ( $\mathbf{P}^*$ ) is a nonclassical resourceful measurement (preparation); (iii) the PVVM  $E^*$  (state  $\varrho^*$ ) nonconvexly overcompletes the nonrefinable effects  $\mathcal{E}_{\text{NR}}$  (states  $\mathcal{S}$ ) into  $\mathcal{E}_{\text{NR}}^*$  ( $\mathcal{S}^*$ ); (iv)  $E^*$  ( $\varrho^*$ ) lies within  $\mathcal{V}^* = \mathcal{V}$  but  $E^* \notin \mathcal{E}$  ( $\varrho^* \notin \mathcal{S}$ ).*

## VII. CONCLUSIONS

We consider the phenomenon of generalized contextuality in general probabilistic theories and show that any GPT satisfying the no-restriction hypothesis is ontologically noncontextual if and only if it is simplicial. We also discuss extensively the case of subGPTs that do not comply with the no-restriction hypothesis. Our results show that any GPT can at most subsume two of the three properties of satisfying the no-restriction hypothesis, ontological noncontextuality, and possessing multiple nonrefinable measurements. Some examples for each possibility already exist, e.g., dropping the no-restriction hypothesis that results in subGPTs as Gaussian quantum mechanics [49], giving up on the ontological noncontextuality as in full quantum theory, or capitulating incompatibility of measurements as in classical mechanics.

A secondary aspect of our work is to provide a new route towards the characterization of the nonclassical power of individual operational elements in information-processing protocols by noting that many nonclassical advantages, though describable by quantum formalism, do not depend on the specifically assumed underlying theory. As an example, in quantum computations we can draw conclusions about their nonclassicality (similarly to the Bell scenario) by merely relying on the classical inputs and outputs and the hardness promises of the computational complexity theory. We thus arrive at two conclusions. First, a nonclassical advantage in a quantum scenario

also implies a nonclassical advantage in any postquantum theory, including those formulated within the GPT framework. Second, the resource formalism suitable for explaining the nonclassical advantages should describe a fundamental theory-independent property, i.e., one that can be certified merely by relying on the measurement statistics. Contextuality is one such property that *ipso facto* forbids a classical description of the processes. Combining these two, in our opinion, the study of contextuality of GPTs and the contextual power of single operational elements is a good candidate for a new approach towards a resource theoretic resolution to the fundamental problem of sufficient resources for quantum computations.

### ACKNOWLEDGMENTS

The author is grateful to John Selby and Rob Spekkens for fruitful discussions. We acknowledge comments by Michael Hall and Markus Müller on an earlier draft of this paper. We also acknowledge support and resources provided by the Royal Commission for the Exhibition of 1851, AQTION project (820495) funded by the European Union Quantum Technology Flagship, and the Sêr SAM Project at Swansea University, an initiative funded through the Welsh Government’s Sêr Cymru II Program (European Regional Development Fund).

*Note added.*—Recently, we became aware of two independent works by Schmid *et al.* [56] and Barnum and Lami [57]. In the former, the concept of *simplex-embeddable* GPTs, which is equivalent to our notion of ontologically noncontextual subGPTs, is introduced. Accordingly, a result similar to Theorem 3 is presented. Here, we further develop the idea by identifying the relationship between the dimensionalities of subGPTs and ontological models. The latter work discusses a result overlapping with Theorem 2. We also became aware of a recent work by Gitton and Woods [58] that report a criterion for noncontextuality with a bound on the size of the ontic spaces. Furthermore, in a recent paper, Wright and Weigert [59] present a theorem similar to our Theorem 1. Both of these researches have been performed independently of ours.

### APPENDIX A: A BRIEF REVIEW OF MEASURES

The way we learn about the notion of “size” in school is a very basic one with value assignments that are real or complex numbers. For example, we learn to calculate the area (or the two-dimensional volume) of a rectangle by multiplying its sides. We are taught later on that the volume under some continuous function can be evaluated by Riemannian integration. This very *sensible* approach is methodological rather than conceptual.

In order to abstractize and henceforth generalize the notion of “size,” we need to pursue the usual mathematical procedure of creating a *mathematical structure*. This route consists of defining some proper sets of objects, some

minimal axioms on those sets that shape the structure, and potentially some suitable functions on those sets. Once this is done to capture “size,” we get to a mathematical structure known as *measure theory*. In the following, we briefly review this procedure for interested physicists.

Suppose a set  $\Omega$  is given. It turns out that “sizes” cannot be defined for one such arbitrary set. Therefore, instead of defining “sizes” for elements of  $\Omega$ , we define them for elements of another set generated by  $\Omega$ , called the  $\sigma$  algebra over  $\Omega$ .

**Definition 1.** *Let  $\Omega$  be a nonempty set. A  $\sigma$  algebra for  $\Omega$  is a subset  $\omega$  of the power set  $2^\Omega$  (i.e., the set of all subsets of  $\Omega$ ) that satisfies the following axioms:*

- (i)  $\omega$  contains  $\Omega$ ;
- (ii) if  $X \in \omega$  then  $\Omega \setminus X \in \omega$ ;
- (iii) for any countable sequence  $\{X_i \in \omega\}_{i \in \mathbb{N}}$  it holds that  $\cup_i X_i \in \omega$ .

**Definition 2.** *The pair  $(\Omega, \omega)$  of an underlying set together with a  $\sigma$  algebra on it is called a measurable space.*

Now, one might be able to develop a sense of why the axioms of the  $\sigma$  algebra make it possible to meaningfully assign “sizes” to its elements. The main idea is that we do have, in a sense, the notions of *complementarity* [axiom (ii)] and *union* [axiom (iii)] of elements. The first one allows us to understand the *relative* “sizes” of different elements, including the relative “size” to the whole collection of elements [axiom (i)]. The second one allows us to grasp the meaning of the “size” of a joint collection of elements. We can now precisely state what we mean by “size.”

**Definition 3.** *Given a measurable space  $(\Omega, \omega)$ , a measure  $\mu$  is a function  $\mu : \omega \rightarrow [0, \infty]$  such that*

- (i)  $\mu(\emptyset) = 0$ ;
- (ii) for any sequence of pairwise disjoint elements  $\{X_i \in \omega\}_{i \in \mathbb{N}}$  it holds that

$$\mu(\cup_i X_i) = \sum_i \mu(X_i). \tag{A1}$$

Here, by pairwise disjoint elements we mean  $X_i \cap X_j = \emptyset$  whenever  $i \neq j$ . Again, a measure has a simple interpretation. For convenience, we need the empty set to have a “size” of zero. This way, we map the identity element of union of sets to the additive identity element of  $\mathbb{R}$ . This is done by axiom (i). Next, we complete the similarity between set union and addition over  $\mathbb{R}$  by condition (ii), which is called the *additivity* property.

It must be now clear that  $\mu$  together with its underlying space deliver all the properties we expect from “size” assignments to objects in a meaningfully general sense.

**Definition 4.** The triplet  $(\Omega, \omega, \mu)$  is called a *measure (or measured) space*.

**Definition 5.** The elements  $X$  of the  $\sigma$  algebra  $\omega$  are called *measurable subsets of  $\Omega$* .

A very useful measure, which is also used within the main text, is Dirac's delta measure. Given the measurable space  $(\Omega, \omega)$ , Dirac delta measure is defined as

$$\delta_X(x) = \begin{cases} 0 & x \notin X \\ 1 & x \in X, \end{cases} \quad (\text{A2})$$

where  $X \in \omega$  is a measurable subset of  $\Omega$  and  $x \in \Omega$ . Dirac delta measure has the ‘‘familiar’’ property that given a measure  $f$  on  $(\Omega, \omega)$  we have

$$\int_{\Omega} f(x) \delta_X(x) dx = f(X). \quad (\text{A3})$$

Consider now the case where  $\Omega$  is the set of all outcomes of some measurement on some physical system. We can then define *events* that are subsets of all possible outcomes. For instance, in flipping a coin the outcomes are ‘‘heads’’ or ‘‘tails’’ and the events are  $\emptyset$ , ‘‘heads,’’ ‘‘tails,’’ ‘‘heads  $\vee$  tails.’’ The event  $\emptyset$  is impossible whereas the event ‘‘heads  $\vee$  tails’’ is certain. It is straightforward to see that the event set can be identified with the power set of the set of measurement outcomes, hence it can be regarded as a  $\sigma$  algebra. As a result, each event is a *measurable subset* of the set of all outcomes. This is, in fact, the case for all measurements on all physical systems, motivating us to define the following measure.

**Definition 6.** Given a measurable space  $(\Omega, \omega)$ , a *probability measure  $p$*  is a measure whose range is the closed interval  $[0, 1]$  and satisfies  $p(\Omega) = 1$ .

Clearly,  $p$  carries the interpretation of a probability distribution over the space of events assigning a ‘‘size’’ to each event that is equivalent to our belief of occurrence of that event.

As mentioned in the main text, indeed, there is no particular reason to restrict the range of measure functions to that of real intervals. Once we replace the range interval  $[0, 1]$  of a probability measure with a subset of a vector space we obtain a probability measure where its values are vectors rather than real numbers, that is a *probability vector-valued measure*, as defined in the main text. Note that the latter holds provided we can identify a fixed identity element ( $U$ ) within the vector space.

## APPENDIX B: PROOF OF THEOREM 1

Following the approach of Refs. [45,46], from the additivity property (iii) and that given any  $E \in \mathcal{E}$ , the effect  $pE$

for any real number  $p \in [0, 1]$  also belongs to  $\mathcal{E}$ , it follows that the map  $q$  is homogeneous over nonnegative rational numbers, i.e.,  $q(mE/n) = mq(E)/n$  for any  $E \in \mathcal{E}$  and  $m, n \in \mathbb{Z}^+$ .

Next, suppose that  $E \in \mathcal{E}$  and  $\alpha, \beta \in [0, 1]$  with  $\alpha < \beta$ . Thus,  $\alpha E$  and  $\beta E$  belong to  $\mathcal{E}$  and  $\alpha E < \beta E$ . It is also clear that  $E' := \beta E - \alpha E = (\beta - \alpha)E$  belongs to  $\mathcal{E}$ . Since  $\beta E = E' + \alpha E$  implies  $q(\beta E) = q(E') + q(\alpha E)$  and  $q(E') \geq 0$  by requirement (i), we have  $q(\alpha E) \leq q(\beta E)$ . That is, the map  $q$  preserves the order of elements within  $\mathcal{E}$ .

Now consider a pair of increasing and decreasing sequences of rational numbers in the  $[0, 1]$  interval,  $(\alpha_i)_i$  and  $(\beta_i)_i$ , respectively, where both converge to the same irrational value  $\gamma \in [0, 1]$ . From the order-preserving property of  $q$  and its homogeneity over rational numbers we obtain  $\alpha_i q(E) = q(\alpha_i E) \leq q(\gamma E) \leq q(\beta_i E) = \beta_i q(E)$ . Then by the pinching theorem we have  $q(\gamma E) = \gamma q(E)$ , that is, the map  $q$  must be linear.

Since  $q$  is a convex-linear functional that is defined on a spanning convex subset ( $\mathcal{E}$ ) of a vector space ( $\mathcal{V}$ ), it can be uniquely extended to a linear functional on the whole vector space ( $\mathcal{V}$ ). Finally, by Riesz's theorem, this linear functional can be written as the inner product  $q(A) = \langle A, B \rangle$  for a unique vector  $B \in \mathcal{V}$ .

The normalization of  $B$  follows simply from requirement (ii) as  $q(U) = \langle U, B \rangle = 1$ .

## APPENDIX C: GPTS AS CONTAINERS FOR SUBGPTS

**Proposition 1.** Any GPT  $\mathcal{T}_{\text{sub}}$  defined on a vector space  $\mathcal{V}$  and identified by the pair  $\mathcal{T}_{\text{sub}} := (\mathcal{E}_{\text{sub}}, \mathcal{S}_{\text{sub}})$  of its physically allowed PVVMs and states that does not satisfy the no-restriction hypothesis is a subtheory of possibly (infinitely) many extended GPTs that do satisfy it.

*Proof.* It follows from violation of the no-restriction hypothesis that  $\mathcal{S}_{\text{sub}} \subset \mathcal{S}$ , where  $\mathcal{S}$  is defined as

$$\mathcal{S} := \{\varrho \in \mathcal{V} | \langle \varrho, E(X) \rangle \geq 0 \forall E(X) \in \mathcal{E}_{\text{sub}}, \langle U, \varrho \rangle = 1\}.$$

Hence,  $\mathcal{T}_{\text{sub}}$  is a subtheory of  $\mathcal{T} := (\mathcal{E}_{\text{sub}}, \mathcal{S})$ , where  $\mathcal{T}$  satisfies the no-restriction hypothesis; see Fig. 2(a). Alternatively, one can fix the state space and define the set of effects as

$$\mathcal{E} := \{E \in \mathcal{V} | \langle \varrho, E \rangle \in [0, 1] \forall \varrho \in \mathcal{S}_{\text{sub}}\}.$$

Clearly this time  $\mathcal{E}_{\text{sub}} \subset \mathcal{E}$  and thus  $\mathcal{T}_{\text{sub}}$  is a subtheory of  $\mathcal{T}' := (\mathcal{E}, \mathcal{S}_{\text{sub}})$ , where  $\mathcal{T}'$  satisfies the no-restriction hypothesis; see Fig. 2(b).

Finally, any GPT  $\mathcal{T}'' := (\mathcal{E}'', \mathcal{S}'')$  such that  $\mathcal{E}_{\text{sub}} \subset \mathcal{E}'' \subseteq \mathcal{E}$ ,  $\mathcal{S}_{\text{sub}} \subset \mathcal{S}'' \subseteq \mathcal{S}$ , and  $\mathcal{E}''$  and  $\mathcal{S}''$  are dual sets, contains  $\mathcal{T}_{\text{sub}}$  as its subtheory; see Fig. 2(c). ■

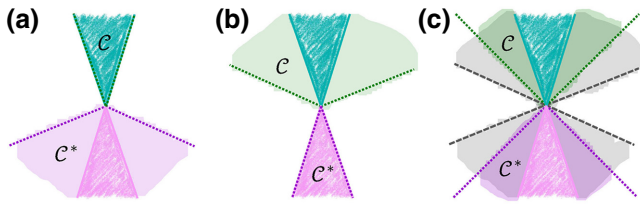


FIG. 2. A heuristic representation of the cones containing three possible GPTs sharing a subGPT. As described in the proof of Proposition 1, (a)  $T = (\mathcal{E}_{\text{sub}}, \mathcal{S})$ , where  $\mathcal{S}_{\text{sub}} \subseteq \mathcal{S}$ ; (b)  $T' = (\mathcal{E}, \mathcal{S}_{\text{sub}})$ , where  $\mathcal{E}_{\text{sub}} \subseteq \mathcal{E}$ ; (c)  $T'' = (\mathcal{E}'', \mathcal{S}'')$ , where  $\mathcal{E}_{\text{sub}} \subseteq \mathcal{E}'' \subseteq \mathcal{E}$  and  $\mathcal{S}_{\text{sub}} \subseteq \mathcal{S}'' \subseteq \mathcal{S}$ . While the subGPT does not satisfy the no-restriction hypothesis, the extended GPTs do satisfy it, i.e., their effect and state spaces are dual to each other. Note that the effect or the state space of all three GPTs can be restricted in such a way to reproduce the given subGPT.

#### APPENDIX D: GENERATORS OF GPT STATE AND EFFECT SPACES

**Lemma 1.** *The pair  $(\mathcal{F}, \mathcal{D})$  mapping a GPT that satisfies the no-restriction hypothesis to a NCOM must be the generating sets of the pair of closed convex sets  $(\mathcal{E}, \mathcal{S})$ . That is,  $\mathcal{E} = \overline{\text{conv}\mathcal{F}}$  and  $\mathcal{S} = \overline{\text{conv}\mathcal{D}}$ .*

*Proof.* In order to have a NCOM, it is first required that  $\zeta_E(X|\lambda) = \langle E(X), D(\lambda) \rangle \geq 0$  for all  $E \in \mathcal{E}$  and all  $\lambda \in \Upsilon$ . Therefore, in view of the no-restriction hypothesis as in the definition of the GPT's state space in Eq. (4) of the main text and that  $\langle U, D(\lambda) \rangle = 1$  for all  $\lambda \in \Upsilon$ , we find that  $D(\lambda) \in \mathcal{S}$ , which implies  $\overline{\text{conv}\mathcal{D}} \subseteq \mathcal{S}$ . Second, by imposing the second requirement of a noncontextual ontological model on Eq. (10), that is,  $\eta_\varrho(\lambda) \geq 0$  for all  $\varrho \in \mathcal{S}$  and all  $\lambda \in \Upsilon$ , we see that  $\mathcal{S} \subseteq \overline{\text{conv}\mathcal{D}(\lambda)}$ . Combining together, it must hold true that  $\overline{\text{conv}\mathcal{D}} = \mathcal{S}$  and thus, the set  $\mathcal{D}$  is a generating set of the GPT's state space  $\mathcal{S}$ .

Similarly, starting from the requirement  $\eta_\varrho(\lambda) = \langle \varrho, F(\lambda) \rangle \geq 0$  for all  $\varrho \in \mathcal{S}$  and all  $\lambda \in \Upsilon$  for a NCOM and noting that the set of effects is dual to the set of states with  $\int_\Upsilon dF(\lambda) = U$ , we infer that  $F(\lambda) \in \mathcal{E}$  and thus,  $\overline{\text{conv}\mathcal{F}} \subseteq \mathcal{E}$ . Next, using the fact that  $\zeta_E(X|\lambda) \geq 0$  for all  $E \in \mathcal{E}$  and all  $\lambda \in \Upsilon$  in Eq. (10) of the main text, that is,

$$\varrho = \int_\Upsilon d\eta_\varrho(\lambda)D(\lambda), \quad \text{and} \quad E(X) = \int_\Upsilon dF(\lambda)\zeta_E(X|\lambda), \quad (\text{D1})$$

we find that  $\mathcal{E} \subseteq \overline{\text{conv}\mathcal{F}}$ . Together, we have  $\mathcal{E} = \overline{\text{conv}\mathcal{F}}$ , meaning that the set  $\mathcal{F}$  is a generating set of the GPT's set of allowed effects  $\mathcal{E}$ . ■

#### APPENDIX E: PROOF OF THEOREM 2

To give the proof of the theorem, we first need to state and prove the following geometrically intuitive lemma.

**Lemma 2.** *Given a nonconvexly overcomplete basis  $\mathcal{A} = \{A_i\}$  of vectors for an ordered vector space  $\mathcal{V}$ , given that elements of  $\mathcal{A}$  belong to the positive pointed generating cone  $\mathcal{C}$ , there exists an element  $C \in \overline{\text{conv}\mathcal{A}}$  such that its convex decomposition into elements of  $\mathcal{A}$  is not unique.*

*Proof of Lemma E.* First, the overcompleteness of  $\mathcal{A}$  means that there exists a vector  $A_J \in \mathcal{A}$  that is linearly dependent on the elements in  $\mathcal{A} \setminus A_J$ . Nonconvexly overcompleteness thus means that  $A_J$  possesses a nonconvex decomposition  $A_J = \sum_{i \neq J} \alpha_i A_i$  in terms of other elements of  $\mathcal{A}$  such that at least one of the expansion coefficients  $\alpha_i$  is negative.

Our proof of the lemma is constructive. We first show that there exists a vector  $B \in \overline{\text{conv}\mathcal{A}}$  such that  $B \notin \overline{\text{conv}\mathcal{A} \setminus A_J}$  and

$$B = \sum_{i \neq J} \beta_i A_i, \quad \sum_{i \neq J} \beta_i = 1. \quad (\text{E1})$$

Considering the vector  $A_J$ , if  $\sum_{i \neq J} \alpha_i > 0$  then we can simply set  $B = A_J / \sum_{i \neq J} \alpha_i$ . Otherwise, bearing in mind that due to being an element of a positive cone all  $\alpha_i$ 's cannot simultaneously be negative, we consider a coefficient  $0 < \alpha^* \in \{\alpha_i\}_{i \neq J}$  that corresponds to the operator  $A^* \in \{A_i\}_{i \neq J}$ . Define the operator  $B(p) := pA_J + (1-p)A^* = p \sum_{i \neq J} \alpha_i A_i + (1-p)A^*$ . Then, for the expansion coefficients of  $B(p)$  it holds true that  $\sum_{i \neq J} \beta_i(p) = p \sum_{i \neq J} \alpha_i + (1-p)$ . By defining  $\bar{\alpha} := |\sum_{i \neq J} \alpha_i|$  we find

$$\bar{p} := p = \begin{cases} \text{any } p \in (0, 1), & \text{if } \bar{\alpha} = 0, \\ \frac{1}{1 + \bar{\alpha}} & \text{otherwise,} \end{cases} \quad (\text{E2})$$

for which  $\sum_{i \neq J} \beta_i(\bar{p}) = 1$ . Evidently,  $B(\bar{p}) \in \overline{\text{conv}\mathcal{A}}$  by construction. However, because there exists at least one  $\beta_i(\bar{p}) = \bar{p}\alpha_i < 0$ , it also holds true that  $B(\bar{p}) \notin \overline{\text{conv}\mathcal{A} \setminus A_J}$ . We thus can set  $B = B(\bar{p})$ .

Given the operator  $B$  with the properties as in Eq. (E1), we consider two sets of indices:  $\mathcal{I}^- := \{i \neq J | \beta_i < 0\}$  and  $\mathcal{I}^+ := \{i \neq J | \beta_i > 0\}$  and define the operator

$$C := \frac{1}{N} (B + \sum_{i \in \mathcal{I}^-} |\beta_i| A_i) = \frac{1}{N} \sum_{i \in \mathcal{I}^+} \beta_i A_i, \quad (\text{E3})$$

in which  $N = \sum_{i \in \mathcal{I}^+} \beta_i$ . We see that, both sides of Eq. (E3) are convex decompositions of  $C$  into elements of  $\mathcal{A}$ , while only the first decomposition contains the operator  $A_J$  (implicit in  $B$ ). ■

*Proof of Theorem 2.* If the points of  $\mathcal{E}_{\text{NR}}$  form a nonconvexly overcomplete basis for  $\mathcal{V}$ , then by Lemma 2 above, there exists a vector (an effect)  $C$  within  $\mathcal{E}$  which is not a coarse-grained effect and yet it possesses a nonunique



decomposition in terms of  $\mathcal{E}_{\text{NR}} \subset \mathcal{F}$ . This means the effect  $C$  possesses a nonunique representation in the ontological model that is not due to coarse graining. As a result, such a GPT does not admit a NCOM. A similar argument holds if the extreme points of  $\mathcal{S}$  form a nonconvexly overcomplete basis. ■

### APPENDIX F: PROOF OF THEOREM 3

The if direction can be shown as follows. All subGPTs  $\mathcal{T}_{\text{sub}}$  of a hypothetical ontologically noncontextual GPT  $\mathcal{T} = (\mathcal{E}, \mathcal{S})$  are ontologically noncontextual provided that  $\dim \mathcal{V}_{\text{sub}} = \dim \mathcal{V}$ . To see this, we note that for any subGPT  $\mathcal{T}_{\text{sub}} = (\mathcal{E}_{\text{sub}}, \mathcal{S}_{\text{sub}})$  it holds true that  $\mathcal{E}_{\text{sub}} \subseteq \mathcal{E}$  and  $\mathcal{S}_{\text{sub}} \subseteq \mathcal{S}$ , where  $\mathcal{E}$  and  $\mathcal{S}$  are the effect and state spaces of the hypothetical *ontologically noncontextual* GPT  $\mathcal{T}$ . Furthermore, by Theorem 2 the hypothetical ontologically noncontextual GPT is simplicial and isomorphic to a hypothetical simplicial ontological model. It is noteworthy that, despite the aforementioned isomorphism, we can safely assume the existence of the hypothetical ontological model without assuming the existence of the hypothetical GPT. This is because the existence of the hypothetical GPT requires the extra assumption that its states and measurements not within the subGPT are physically realizable [60], whereas the existence of the hypothetical ontological model does not make such an assumption, i.e., it does not presume that its states and measurements not within the subontological model associated to the subGPT are physically realizable.

Now, for all elements  $E(X) \in \mathcal{E}_{\text{sub}}$  and  $\varrho \in \mathcal{S}_{\text{sub}}$  one can simply use the elements of the hypothetical ontological model  $\zeta_E(X|\lambda)$  and  $\eta_\varrho(\lambda)$  that are the same ones assigned to the states and effects of the hypothetical GPT  $\mathcal{T} = (\mathcal{E}, \mathcal{S})$ . These assignments trivially reproduce the statistics of the subGPT, meaning that there exists (at least) one subontological model to which the subGPT can be associated. There is, however, a subtle point: the physically allowed measurements are only those defined by the subGPT as  $\mathcal{E}_{\text{sub}}$ . As a result, if  $\dim \mathcal{V}_{\text{sub}} < \dim \mathcal{V}$ , then for every state of the subGPT there will exist multiple ontic states with components in  $\mathcal{V}_{\text{sub}}^\perp \subset \mathcal{V}$  that generate exactly the same statistics under the physically possible measurements, hence, the subontological model associated with the subGPT will not be unique. In other words,  $\mathbf{P}_1 \cong \mathbf{P}_2 \not\Rightarrow \mu_{\mathbf{P}_1} = \mu_{\mathbf{P}_2}$  and thus, the ontological model will not be satisfying the broad noncontextuality assumption (see Sec. 2 below for an explicit example).

The only if direction can be shown as follows. Assume that the subtheory  $\mathcal{T}_{\text{sub}} = (\mathcal{E}_{\text{sub}}, \mathcal{S}_{\text{sub}})$  admits a NCOM. This means that there exist bijective convex linear maps  $\eta_{\text{sub}}$  and  $\zeta_{\text{sub}}$  from state and effect spaces of the subtheory to the ontic state and indicator functions over some ontic variable space  $\Upsilon$ . We note that we can always assume that the effect and state spaces of the subGPT span its underlying vector space  $\mathcal{V}_{\text{sub}}$ . As a result,  $\eta_{\text{sub}}$  and  $\zeta_{\text{sub}}$  can be uniquely

extended to bijective maps  $\eta$  and  $\zeta$  over the whole  $\mathcal{V}_{\text{sub}}$ . We also know that, due to being probability measures on  $\Upsilon$ , the ontic state and indicator function spaces of the NCOM can be enlarged to ontic state and measurement spaces whose extreme points are Dirac delta measures. Indeed, the extended ontic spaces are dual to each other and satisfy the no-restriction hypothesis, identifying a hypothetical simplicial ontological model. In the next step, we simply apply  $\eta^{-1}$  and  $\zeta^{-1}$ , the inverses of the extended maps  $\eta$  and  $\zeta$ , to the extended ontic spaces to obtain extended state and effect spaces  $\mathcal{S} \supseteq \mathcal{S}_{\text{sub}}$  and  $\mathcal{E} \supseteq \mathcal{E}_{\text{sub}}$  in  $\mathcal{V}$ . It is immediate that  $\dim \mathcal{V} = \dim \mathcal{V}_{\text{sub}}$ , and  $\mathcal{E}$  and  $\mathcal{S}$  are dual thus the GPT satisfies the no-restriction hypothesis that identifies the desired hypothetical GPT. Moreover,  $\mathcal{T} = (\mathcal{E}, \mathcal{S})$  is ontologically noncontextual by construction. Finally,  $\mathcal{T}_{\text{sub}}$  is a subtheory of  $\mathcal{T}$ . We emphasize that the GPT  $\mathcal{T}$  constructed in this way is only a hypothetical theory in the sense that the physically allowed states and measurements are only those described by the subGPT  $\mathcal{T}_{\text{sub}}$ . Hence, there is no assumption that a parent simplicial GPT *physically* exists.

### APPENDIX G: TWO EXAMPLES

#### 1. Spekkens' toy theory

As an interesting example of the application of Theorems 2 and 3, we analyze the noncontextuality of Spekkens' toy theory [53].

In this model, there exists an elementary system whose pure ontic states are denoted by “1,” “2,” “3,” and “4” that can be represented by column vectors

$$\eta_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \eta_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad \eta_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad \eta_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad (\text{G1})$$

respectively, i.e. a simplex in  $\mathbb{R}^4$ ; see also Fig. 3. The measurements on the system can be considered as questions about the state of the system. Of particular interest are the two outcome measurements that are *reproducible*, or *repeatable* [52], in the sense that if repeated upon a system in a pure state they always output the same outcome. The set of all such measurements define the sharp effects of the dichotomic measurements, and they are given by “is the state of the system 1 or not?,” “is the state of the system 2 or not?,” “is the state of the system 3 or not?,” and “is the state of the system 4 or not?” Naturally, these measurements are coarse-grained versions of a single four-outcome nonrefinable measurement that can also be mapped to a

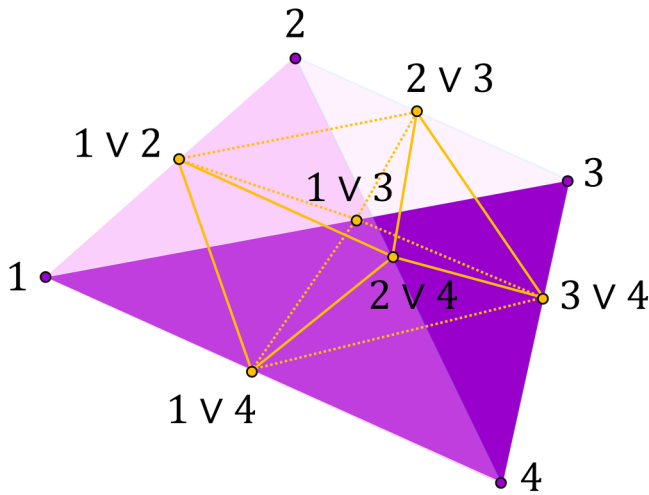


FIG. 3. Three-dimensional representation of the (ontic) state and effect spaces of the GPT underlying Spekkens’ toy theory. This is a simplex whose vertices correspond to vectors spanning  $\mathbb{R}^4$ . The (epistemic) state and effect spaces of the subGPT (i.e., Spekkens’ toy theory) is obtained by imposing the knowledge balance principle, which are given by octahedra inside the GPT’s state and effect spaces. Clearly, while the state and effect spaces of the GPT are dual to each other and satisfy the no-restriction hypothesis, the state and effect spaces of the subGPT are not duals and do not satisfy this hypothesis. Nevertheless, Spekkens’ toy theory is ontologically noncontextual according to Theorem 3, because it is a subtheory of an ontologically noncontextual GPT.

simplex in  $\mathbb{R}^4$ . This PVVM is given by

$$\zeta_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}^\top, \quad \zeta_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}^\top, \quad \zeta_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}^\top, \quad \zeta_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}^\top, \quad (\text{G2})$$

with the unit element  $\zeta_U = (1, 1, 1, 1)$ . Note that the set of all PVVMs, which is dual to the state space forms a hypercube. As such, according to Theorem 2, we have a four-dimensional classical GPT for the elementary system that satisfies the no-restriction hypothesis.

Given the elementary system and the GPT for it, one can define a *canonical measurement set* to be a minimal set of possibly coarse-grained measurements that fully determine the state of the system. For the elementary system above, one such set, for example, is given by {“is the state of the system 1 OR 2, or not?”, “is the state of the system 1 OR 3, or not?”}. At this point, it is possible to define the *amount of knowledge* to be the maximum number of questions for which the answer is known, varied over all possible canonical measurement sets.

Given a measure of knowledge, an *epistemic* restriction can be imposed on the theory, which results in Spekkens’ toy theory.

*The knowledge balance principle* [53]: If one has the maximal knowledge, then for every system, at every time, the amount of knowledge one possesses about the ontic state of the system at that time must equal the amount of knowledge one lacks.

After enforcing the knowledge balance principle on the original theory, a subtheory is obtained; the pure states of which are given as “1 ∨ 2,” “1 ∨ 3,” “1 ∨ 4,” “2 ∨ 3,” “2 ∨ 4,” and “3 ∨ 4” [53]. Here, ∨ denotes the disjunction or “OR” operator. These states can be represented by column vectors

$$\eta_5 = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 0 \\ 0 \end{pmatrix}, \quad \eta_6 = \begin{pmatrix} 0 \\ \frac{1}{2} \\ \frac{1}{2} \\ 0 \end{pmatrix}, \quad \eta_7 = \begin{pmatrix} \frac{1}{2} \\ 0 \\ 1 \\ \frac{1}{2} \end{pmatrix}, \quad \eta_8 = \begin{pmatrix} 0 \\ \frac{1}{2} \\ 0 \\ \frac{1}{2} \end{pmatrix},$$

$$\eta_9 = \begin{pmatrix} 0 \\ \frac{1}{2} \\ \frac{1}{2} \\ 0 \end{pmatrix}, \quad \eta_{10} = \begin{pmatrix} \frac{1}{2} \\ 0 \\ 0 \\ \frac{1}{2} \end{pmatrix}, \quad (\text{G3})$$

respectively, which form an octahedron that is circumscribed by the GPT’s state space [61]; see Fig. 3. The set of reproducible and nonrefinable PVVMs also reduces to {1 ∨ 2, 3 ∨ 4}, {1 ∨ 3, 2 ∨ 4}, and {1 ∨ 4, 2 ∨ 3}, that can be represented by

$$\zeta_5 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}^\top, \quad \zeta_6 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}^\top, \quad \zeta_7 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}^\top,$$

$$\zeta_8 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}^\top, \quad \zeta_9 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}^\top, \quad \zeta_{10} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}^\top, \quad (\text{G4})$$

respectively, which also form an octahedron inside the GPT’s effect space; see Fig. 3. By the restriction enacted, the effect and state spaces of the subGPT do not satisfy the no-restriction hypothesis anymore. It is clear that  $\dim \mathcal{V}_{\text{Spekkens}} = 4$  equals the dimensionality of the container GPT. A second way to check this is by considering

the *probability table* of the subGPT, given as

$$\begin{array}{c|cccccc}
 T_{\text{Spekkens}} & E(5) & E(6) & E(7) & E(8) & E(9) & E(10) \\
 \hline
 \mathbf{P}_5 & 1 & 0 & 1/2 & 1/2 & 1/2 & 1/2 \\
 \mathbf{P}_6 & 0 & 1 & 1/2 & 1/2 & 1/2 & 1/2 \\
 = \mathbf{P}_7 & 1/2 & 1/2 & 1 & 0 & 1/2 & 1/2 \\
 \mathbf{P}_8 & 1/2 & 1/2 & 0 & 1 & 1/2 & 1/2 \\
 \mathbf{P}_9 & 1/2 & 1/2 & 1/2 & 1/2 & 1 & 0 \\
 \mathbf{P}_{10} & 1/2 & 1/2 & 1/2 & 1/2 & 0 & 1
 \end{array} \quad , \quad (\text{G5})$$

and checking its rank. In this case,  $\text{rank} T_{\text{Spekkens}} = 4$ . Given a probability table for sets of pure preparations and nonrefinable measurements, it is known that any model for that table must satisfy  $\dim \mathcal{V}_{\text{model}} \geq \text{rank} T$  [14], where for the Spekkens' toy model the equality can be satisfied. Consequently, by Theorem 3, Spekkens' subGPT is a proper ontologically noncontextual subtheory.

## 2. Stabilizer rebit theory

Our second example is closely related to but significantly different from the previous one. The subtheory of stabilizer rebit is a model in which the nonrefinable effects and pure states of a qubit system is restricted to  $E(1), \mathbf{P}_1 \mapsto |0\rangle$ ,  $E(2), \mathbf{P}_2 \mapsto |1\rangle$ ,  $E(3), \mathbf{P}_3 \mapsto |+\rangle$ ,  $E(4), \mathbf{P}_4 \mapsto |-\rangle$ . These produce a table of probabilities for *all in principle possible* prepare and measure experiments on a rebit given by

$$\begin{array}{c|cccc}
 T_{\text{rebit}} & E(1) & E(2) & E(3) & E(4) \\
 \hline
 \mathbf{P}_1 & 1 & 0 & 1/2 & 1/2 \\
 \mathbf{P}_2 & 0 & 1 & 1/2 & 1/2 \\
 \mathbf{P}_3 & 1/2 & 1/2 & 1 & 0 \\
 \mathbf{P}_4 & 1/2 & 1/2 & 0 & 1
 \end{array} \quad . \quad (\text{G6})$$

It is evident that  $T_{\text{rebit}}$  is a subtable of  $T_{\text{Spekkens}}$  in Eq. (G5).

Suppose that we want to construct a GPT for this table. The first question is what the dimensionality of the employed vector space should be. To answer this question, we note that  $\text{rank} T_{\text{rebit}} = 3$  and thus, regardless of the model constructed being a GPT or an ontological model,  $\dim \mathcal{V}_{\text{model}} \geq 3$ .

A GPT on a three-dimensional space reproducing  $T_1$  can be given by the states

$$\begin{aligned}
 s_1 &= \begin{pmatrix} 1 \\ \frac{1}{2} \\ 0 \\ \frac{1}{2} \end{pmatrix}, & s_2 &= \begin{pmatrix} -\frac{1}{2} \\ 0 \\ 1 \\ \frac{1}{2} \end{pmatrix}, & s_3 &= \begin{pmatrix} 0 \\ \frac{1}{2} \\ 1 \\ -\frac{1}{2} \end{pmatrix}, \\
 s_4 &= \begin{pmatrix} 0 \\ -\frac{1}{2} \\ 1 \\ \frac{1}{2} \end{pmatrix},
 \end{aligned} \quad (\text{G7})$$

and the effects

$$\begin{aligned}
 e_1 &= \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}^\top, & e_2 &= \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}^\top, & e_3 &= \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}^\top, \\
 e_4 &= \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}^\top,
 \end{aligned} \quad (\text{G8})$$

with the unit element  $e_U = (0, 0, 2)$ .

We note that the given GPT does not satisfy the no-restriction hypothesis, i.e., it is a subGPT. One GPT containing this subGPT, for instance, is given by the state space generated by the vectors

$$\begin{aligned}
 s_5 &= \begin{pmatrix} 1 \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}, & s_6 &= \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}, & s_7 &= \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}, \\
 s_8 &= \begin{pmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ 1 \\ \frac{1}{2} \end{pmatrix}.
 \end{aligned} \quad (\text{G9})$$

These states form the extreme points of the dual cone to the effect space defined by  $e_1, e_2, e_3$ , and  $e_4$  with a positive inner product with all the given effects and thus, if the no-restriction hypothesis holds, they must be legitimate states. It is clear that the states  $s_1, s_2, s_3$ , and  $s_4$  can be written as convex combinations of  $s_5, s_6, s_7$ , and  $s_8$ . For instance,  $s_1 = (s_5 + s_7)/2$ . From Theorem 2 follows that this GPT is ontologically contextual.

Given the four pure states and the four nonrefinable effects of table  $T_{\text{ssrebit}}$  in Eq. (G6), no ontological model that respects the broad noncontextuality and reproduces the desired probabilities exists for  $\text{card } \Upsilon = \dim \mathcal{V}_{\text{ontic}} = 3$ . Hence, the subGPT is ontologically contextual for

card  $\Upsilon = \dim \mathcal{V}_{\text{ontic}} = 3$ . We now show that increasing dimensionality does not resolve this issue.

Quite easily, it is possible to construct an ontological model for the above subGPT by enlarging the ontic space such that  $\text{card } \Upsilon = \dim \mathcal{V}_{\text{ontic}} = 4$ . We can represent preparations and measurement effects as ontic states and indicator functions

$$\eta_5 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \eta_6 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad \eta_7 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad \eta_8 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \quad (\text{G10})$$

and

$$\zeta_5 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}^T, \quad \zeta_6 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}^T, \quad \zeta_7 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}^T, \quad \zeta_8 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}^T, \quad (\text{G11})$$

respectively. This is isomorphic to the GPT containing Spekkens' toy theory. There are various ways in which we can construct a subontological model reproducing  $T_{\text{rebit}}$ , one of which is via the states and indicator functions

$$\eta_1 = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ 0 \end{pmatrix}, \quad \eta_2 = \begin{pmatrix} 0 \\ 0 \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}, \quad \eta_3 = \begin{pmatrix} \frac{1}{2} \\ \frac{2}{0} \\ \frac{1}{2} \\ \frac{2}{0} \end{pmatrix}, \quad \eta_4 = \begin{pmatrix} 0 \\ \frac{1}{2} \\ \frac{2}{0} \\ \frac{1}{2} \end{pmatrix} \quad (\text{G12})$$

and

$$\zeta_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}^T, \quad \zeta_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}^T, \quad \zeta_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}^T, \quad \zeta_4 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}^T. \quad (\text{G13})$$

The most important point here is that, despite being an ontological model, this model does not satisfy broad noncontextuality when noting that we are operationally restricted to the statistics given by the subGPT. In other words, assuming the subGPT  $T_{\text{rebit}}$  (or in general, any GPT) implicitly implies that no preparation and measurement beyond what is predicted by the subGPT exist. As a

result, we can readily verify that the vector

$$\eta^* = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 0 \\ \frac{1}{2} \end{pmatrix}, \quad (\text{G14})$$

produces the same statistics as the maximally mixed rebit state

$$\eta_{\text{MM}} = \begin{pmatrix} \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \end{pmatrix}, \quad (\text{G15})$$

for every *physically legitimate* ontic effect generated by  $\{\zeta_5, \zeta_6, \zeta_7, \zeta_8\}$  of Eq. (G13).

Note that,  $\eta^*$  does not belong to the state space of the subontological model corresponding to the stabilizer rebit subGPT  $\mathcal{S}_{\text{rebit}}$  but it is an ontic state of the full ontological model, i.e., it is a convex combination of states in Eq. (G10). As such, given the preparation procedure  $\mathbf{P}_{\text{MM}}$  that generates the table

$$T_{\text{MM}} = \frac{\quad}{\mathbf{P}_{\text{MM}}} \begin{array}{c|cccc} & E(1) & E(2) & E(3) & E(4) \\ \hline & 1/2 & 1/2 & 1/2 & 1/2 \end{array}, \quad (\text{G16})$$

we have two possible ontic assignments:

$$\mathbf{P}_{\text{MM}} \mapsto \eta_{\text{MM}} \quad \text{or} \quad \mathbf{P}_{\text{MM}} \mapsto \eta^*. \quad (\text{G17})$$

Therefore, even though  $\eta_{\text{MM}}$  and  $\eta^*$  are distinct ontic functions with regards to the full ontic measurements of Eq. (G11), by the implicit assumption of the stabilizer rebit subGPT, it is operationally impossible to distinguish them. As a result, Eq. (G17) is in clear contradiction with broad noncontextuality assumption for NCOMs [Eq. (5)] and the fact that each preparation equivalence class must be represented by a *unique* ontic state that is completely determined by the observable statistics.

It is worth pointing out that similar arguments can also be made by exchanging the roles of ontic states and indicator functions.

## APPENDIX H: IS THERE A CONFLICT?

An interesting lesson we learn from the example of stabilizer rebit subtheory is that restricting classical theories [in this case the ontologically noncontextual GPT given by Eqs. (G10) and (G11)] in certain ways may also give rise to ontologically contextual subtheories. This faces us with

the dilemma of choosing between (i) assuming all subtheories of classical theories are classical and thus ontological contextuality is not a purely nonclassical effect, or (ii) ontological noncontextuality is a purely nonclassical effect and thus subtheories of classical theories can also be nonclassical. Our personal preference is the latter, because the assumption of (i) cannot be justified. In particular, there is no reason to call a weird-looking world wherein our empirical observations are restricted in such a way that certain statistics have multiple undeterminable ontic explanations classical, noting that any such explanation is different from our current classical perception of the world.

One might also be concerned that the claim “stabilizer rebit subGPT is ontologically contextual” conflicts with other results, e.g., Refs. [7] and [16], or that this subGPT can be classically *simulated*. We, however, emphasize that there is no conflicts with either of these results for the following reasons.

In Refs. [7,16], the idea is to build a NCOM for a given *table of probabilities* rather than a subGPT. There is a significant difference between the two, namely that, once a subGPT (like the stabilizer rebit theory) is given it is assumed that

no observable statistics exist beyond those predicted by the (sub)theory.

Therefore, if a NCOM exists in larger spaces, the extra statistics predicted by the NCOM is empirically unobservable leading to the previously discussed contradiction. In sharp contrast, given merely a set of experimental data there is *no* assumption that the given data is *all one can in principle measure for the physical system*. Hence, once a NCOM for the data is built (possibly in larger dimensions) it is implicit that the extra statistics predicted by the model but not included in the given table are indeed empirically observable. Hence, the difference between the two scenarios is in assuming or not assuming empirical observability of the statistics predicted by the NCOM beyond the given table or the subGPT, respectively.

A similar argument holds when we speak of classical simulatability, wherein, again, there is no assumption of physical possibility or impossibility of certain statistics. Therefore, as an example, the stabilizer rebit subGPT being ontologically contextual does *not* imply it being classically nonsimulatable. This highlights the conceptual difference between “classicality” and “classical simulatability,” which was also pointed out by Meyer in the context of quantum games in his reply to van Enk [62]:

“Clearly, the expansion of Q’s pure strategy set [4] which enables him to win every game can be realized in either a quantum or a classical system, but to argue that ‘A single qubit is not a truly quantum system’ because it can be ‘mocked up by a classical hidden-variable model’ [2] is, as Heisenberg put it, to ‘attempt to put new wine into old bottles. Such attempts are always distressing, for they mislead

us into continually occupying ourselves with the inevitable cracks in the old bottles, instead of rejoicing over the new wine’ [5].”

Our position of classicality identified by ontological noncontextuality requires, in addition to classical simulatability, that the (sub)GPT is not underdetermined in the sense that there are no *in principle inaccessible* variables present at the ontological level. The latter requirement follows the spirit of Leibniz’s methodological principle, as Spekkens [54] puts it,

“If an ontological theory implies the existence of two scenarios that are empirically indistinguishable in principle but ontologically distinct (where both the indistinguishability and distinctness are evaluated by the lights of the theory in question), then the ontological theory should be rejected and replaced with one relative to which the two scenarios are ontologically identical.”

The principle thus asks for ontologically identical descriptions of facts that cannot be empirically distinguished. It is clear that, in the example of rebit stabilizer subGPT, replacing  $\eta_{MM}$  with  $\eta^*$  results in two ontologically distinct accounts for the same subGPT (i.e., empirical) description. Hence, the rebit stabilizer subGPT does not allow for an ontological model satisfying the Leibniz’s methodological principle and, by our definition, it is a nonclassical subGPT.

#### APPENDIX I: PROOF OF THEOREM 4

That (i) if and only if (ii) follows from the definition of nonclassical resources given above. That (i) if and only if (iii) follows from Theorem 2 and that  $\mathcal{E}_{NR}$  (pure states) forms a complete basis for  $\mathcal{V}$ . That (iii) if and only if (iv) follows from the assumption that  $\mathcal{E}_{NR}$  (pure states) form a complete basis for  $\mathcal{V}$ , hence adding an extra nonrefinable extreme point overcompletes it if  $\mathcal{V}^* = \mathcal{V}$ . Conversely, it is trivial that if an overcompleting element lies within  $\mathcal{E}$ , it is either refinable or a convex combination of the extreme points.

- 
- [1] S. Kochen and E. P. Specker, in *The Logico-Algebraic Approach to Quantum Mechanics: Volume I: Historical Evolution*, edited by C. A. Hooker (Springer Netherlands, Dordrecht, 1975), p. 293.
  - [2] A. Cabello, S. Severini, and A. Winter, (Non-)Contextuality of Physical Theories as an Axiom, [arXiv:1010.2163](https://arxiv.org/abs/1010.2163) (2010).
  - [3] S. Abramsky and A. Brandenburger, The sheaf-theoretic structure of non-locality and contextuality, *New J. Phys.* **13**, 113036 (2011).
  - [4] A. Acín, T. Fritz, A. Leverrier, and A. B. Sainz, A combinatorial approach to nonlocality and contextuality, *Commun. Math. Phys.* **334**, 533 (2015).

- [5] R. W. Spekkens, Contextuality for preparations, transformations, and unsharp measurements, *Phys. Rev. A* **71**, 052108 (2005).
- [6] R. W. Spekkens, Negativity and Contextuality are Equivalent Notions of Nonclassicality, *Phys. Rev. Lett.* **101**, 020401 (2008).
- [7] D. Schmid and R. W. Spekkens, Contextual Advantage for State Discrimination, *Phys. Rev. X* **8**, 011015 (2018).
- [8] L. Hardy, Quantum Theory From Five Reasonable Axioms, [arXiv:0101012](https://arxiv.org/abs/0101012) [quant-ph] (2001).
- [9] J. Barrett, Information processing in generalized probabilistic theories, *Phys. Rev. A* **75**, 032304 (2007).
- [10] G. Chiribella, G. M. D’Ariano, and P. Perinotti, Probabilistic theories with purification, *Phys. Rev. A* **81**, 062348 (2010).
- [11] G. Chiribella, G. M. D’Ariano, and P. Perinotti, Informational derivation of quantum theory, *Phys. Rev. A* **84**, 012311 (2011).
- [12] L. Hardy, Reformulating and Reconstructing Quantum Theory, [arXiv:1104.2066](https://arxiv.org/abs/1104.2066) (2011).
- [13] L. Masanes and M. P. Müller, A derivation of quantum theory from physical requirements, *New J. Phys.* **13**, 063001 (2011).
- [14] P. Janotta and H. Hinrichsen, Generalized probability theories: What determines the structure of quantum theory? *J. Phys. A: Math. Theor.* **47**, 323001 (2014).
- [15] R. Kunjwal and R. W. Spekkens, From the Kochen-Specker Theorem to Noncontextuality Inequalities Without Assuming Determinism, *Phys. Rev. Lett.* **115**, 110403 (2015).
- [16] D. Schmid, R. W. Spekkens, and E. Wolfe, All the non-contextuality inequalities for arbitrary prepare-and-measure experiments with respect to any fixed set of operational equivalences, *Phys. Rev. A* **97**, 062103 (2018).
- [17] A. Mari and J. Eisert, Positive Wigner Functions Render Classical Simulation of Quantum Computation Efficient, *Phys. Rev. Lett.* **109**, 230503 (2012).
- [18] V. Veitch, N. Wiebe, C. Ferrie, and J. Emerson, Efficient simulation scheme for a class of quantum optics experiments with non-negative Wigner representation, *New J. Phys.* **15**, 013037 (2013).
- [19] V. Veitch, S. A. Hamed Mousavian, D. Gottesman, and J. Emerson, The resource theory of stabilizer quantum computation, *New J. Phys.* **16**, 013009 (2014).
- [20] R. Raussendorf, J. Bermejo-Vega, E. Tyhurst, C. Okay, and M. Zurek, Phase space simulation method for quantum computation with magic states on qubits, [arXiv:1905.05374](https://arxiv.org/abs/1905.05374) (2019).
- [21] D. Gottesman and I. L. Chuang, Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations, *Nature* **402**, 390 (1999).
- [22] S. Bravyi and A. Kitaev, Universal quantum computation with ideal Clifford gates and noisy ancillas, *Phys. Rev. A* **71**, 022316 (2005).
- [23] E. Knill, Quantum computing with realistically noisy devices, *Nature* **434**, 39 (2005).
- [24] M. Howard, J. Wallman, V. Veitch, and J. Emerson, Contextuality supplies the ‘magic’ for quantum computation, *Nature* **510**, 351 (2014).
- [25] E. Chitambar and G. Gour, Quantum resource theories, *Rev. Mod. Phys.* **91**, 025001 (2019).
- [26] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Quantum entanglement, *Rev. Mod. Phys.* **81**, 865 (2009).
- [27] F. G. S. L. Brandão, M. Horodecki, J. Oppenheim, J. M. Renes, and R. W. Spekkens, Resource Theory of Quantum States out of Thermal Equilibrium, *Phys. Rev. Lett.* **111**, 250404 (2013).
- [28] M. Horodecki and J. Oppenheim, Fundamental limitations for quantum and nanoscale thermodynamics, *Nat. Commun.* **4**, 2059 (2013).
- [29] F. Brandão, M. Horodecki, N. Ng, J. Oppenheim, and S. Wehner, The second laws of quantum thermodynamics, *Proc. Natl. Acad. Sci.* **112**, 3275 (2015).
- [30] P. Faist, F. Dupuis, J. Oppenheim, and R. Renner, The minimal work cost of information processing, *Nat. Commun.* **6**, 7669 (2015).
- [31] M. Lostaglio, K. Korzekwa, D. Jennings, and T. Rudolph, Quantum Coherence, Time-Translation Symmetry, and Thermodynamics, *Phys. Rev. X* **5**, 021001 (2015).
- [32] G. Gour, M. P. Müller, V. Narasimhachar, R. W. Spekkens, and N. Y. Halpern, The resource theory of informational nonequilibrium in thermodynamic, *Phys. Rep.* **583**, 1 (2015).
- [33] J. Aberg, Quantifying superposition, [arXiv:0612146](https://arxiv.org/abs/0612146) (2006).
- [34] T. Baumgratz, M. Cramer, and M. B. Plenio, Quantifying Coherence, *Phys. Rev. Lett.* **113**, 140401 (2014).
- [35] A. Winter and D. Yang, Operational Resource Theory of Coherence, *Phys. Rev. Lett.* **116**, 120404 (2016).
- [36] A. Streltsov, G. Adesso, and M. B. Plenio, Colloquium: Quantum coherence as a resource, *Rev. Mod. Phys.* **89**, 041003 (2017).
- [37] G. Gour and R. W. Spekkens, The resource theory of quantum reference frames: Manipulations and monotones, *New J. Phys.* **10**, 033023 (2008).
- [38] G. Gour, I. Marvian, and R. W. Spekkens, Measuring the quality of a quantum reference frame: The relative entropy of frameness, *Phys. Rev. A* **80**, 012307 (2009).
- [39] M. Piani, M. Cianciaruso, T. R. Bromley, C. Napoli, N. Johnston, and G. Adesso, Robustness of asymmetry and coherence of quantum states, *Phys. Rev. A* **93**, 042107 (2016).
- [40] Quantum non-Markovianity: Characterization, quantification and detection, *Rep. Prog. Phys.* **77**, 094001 (2014).
- [41] Quantifying spatial correlations of general quantum dynamics, *New J. Phys.* **17**, 062001 (2015).
- [42] A. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (Edizioni della Normale, Pisa, 2011).
- [43] P. Busch, M. Grabowski, and P. J. Lahti, *Operational Quantum Physics* (Springer, Berlin, Heidelberg, 1995).
- [44] A. M. Gleason, Measures on the closed subspaces of a Hilbert space, *J. Math. Mech.* **6**, 885 (1957).
- [45] P. Busch, Quantum States and Generalized Observables: A Simple Proof of Gleason’s Theorem, *Phys. Rev. Lett.* **91**, 120403 (2003).
- [46] C. M. Caves, C. A. Fuchs, K. K. Manne, and J. M. Renes, Gleason-type derivations of the quantum probability rule for generalized measurements, *Found. Phys.* **34**, 193 (2004).
- [47] H. Barnum, J. Barrett, L. O. Clark, M. Leifer, R. Spekkens, N. Stepanik, A. Wilce, and R. Wilke, Entropy and

- information causality in general probabilistic theories, *New J. Phys.* **12**, 033024 (2010).
- [48] P. Janotta and R. Lal, Generalized probabilistic theories without the no-restriction hypothesis, *Phys. Rev. A* **87**, 052131 (2013).
- [49] S. D. Bartlett, T. Rudolph, and R. W. Spekkens, Reconstruction of Gaussian quantum mechanics from Liouville mechanics with an epistemic restriction, *Phys. Rev. A* **86**, 012103 (2012).
- [50] C. Ferrie and J. Emerson, Framed Hilbert space: Hanging the quasi-probability pictures of quantum theory, *New J. Phys.* **11**, 063040 (2009).
- [51] C. Ferrie and J. Emerson, Frame representations of quantum mechanics and the necessity of negativity in quasi-probability representations, *J. Phys. A Math. Theor.* **41**, 352001 (2008).
- [52] G. Chiribella and X. Yuan, Measurement sharpness cuts nonlocality and contextuality in every physical theory, [arXiv:1404.3348v2](https://arxiv.org/abs/1404.3348v2) (2014).
- [53] R. W. Spekkens, Evidence for the epistemic view of quantum states: A toy theory, *Phys. Rev. A* **75**, 032110 (2007).
- [54] R. W. Spekkens, The ontological identity of empirical indiscernibles: Leibniz's methodological principle and its significance in the work of Einstein, [arXiv:1909.04628](https://arxiv.org/abs/1909.04628) [physics.hist-ph] (2019).
- [55] M. D. Mazurek, M. F. Pusey, R. Kunjwal, K. J. Resch, and R. W. Spekkens, An experimental test of noncontextuality without unphysical idealizations, *Nat. Commun.* **7**, ncomms11780 (2016).
- [56] D. Schmid, J. Selby, E. Wolfe, R. Kunjwal, and R. W. Spekkens, The characterization of noncontextuality in the framework of generalized probabilistic theories, [arXiv:1911.10386](https://arxiv.org/abs/1911.10386) [quant-ph] (2019).
- [57] H. Barnum and L. Lami, Classical embeddability of general probabilistic theories, Forthcoming.
- [58] V. Gitton and M. P. Woods, Solvable Criterion for the Contextuality of any Prepare-and-Measure Scenario, [arXiv:2003.06426](https://arxiv.org/abs/2003.06426) [quant-ph] (2021).
- [59] V. J. Wright, and S. Weigert, General Probabilistic Theories with a Gleason-type Theorem, [arXiv:2005.14166](https://arxiv.org/abs/2005.14166) [quant-ph] (2020).
- [60] Note that, the physically allowed states and measurements are only those defined by the subGPT.
- [61] It is worth emphasizing that, despite the representation given here, Spekkens' toy theory does not allow for all possible convex combinations of its pure states [53]. However, this fact is irrelevant to our analysis.
- [62] D. A. Meyer, Meyer Replies, *Phys. Rev. Lett.* **84**, 790 (2000).