Contextuality of general probabilistic theories and the power of a single resource

Farid Shahandeh$^1$,

$^1$Department of Physics, Swansea University, Singleton Park, Swansea SA2 8PP, United Kingdom

Generalized contextuality refers to our inability of explaining measurement statistics using a context-independent probabilistic model. On the other hand, measurement statistics can also be modeled most generally within the framework of general probabilistic theories. Here, starting from a construction of general probabilistic theories based on a Gleason-type theorem, we show that for any such theory 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. In other words, requiring any two of these properties for a general probabilistic theory implies the violation of the third one. We extensively discuss general probabilistic theories for which (ii) and (iii) are met while (i) is violated, exploring the role of the no-restriction hypothesis as a sufficient condition that enforces ontological contextuality on nonsimplicial general probabilistic theories. 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.

I. INTRODUCTION

Contextuality is a hallmark of nonclassicality [1, 2]. Within the operational approach to physical theories, where the primitive elements are laboratory instructions for preparations and measurements of systems, a generalized notion of contextuality refers to a no-go theorem that dismisses all probabilistic accounts for measured statistics that do not rely upon the “contexts” of preparations and measurements [3]. Classical theories can minimally be regarded as noncontextual probabilistic theories and thus, it is plausible to define nonclassicality as generalized contextuality [4, 5].

Naturally, given collections of preparations and measurements a fundamental question arises as to whether their resulting statistics can be explained using a classical (hence ontologically noncontextual) model. Recently, Kunjwal and Spekkens [6] and Schmid et al [7] provided an answer by deriving noncontextuality inequalities for detection of possible nonclassical statistics. In this paper, we take a different perspective in answering this question: besides ontological models, the operational approach can also be abstracted in a broad sense using the framework of general probabilistic theories (GPTs), containing quantum theory as a special case [8–14]. A nonclassical set of data thus means that none of the potential GPT descriptions can be translated into a noncontextual probabilistic model. We determine which GPTs allow and which ones disallow noncontextual ontological models. More specifically, we prove a Gleason-type theorem for GPTs and show that, in finite dimensions, any GPT that is built upon this theorem and satisfies the no-restriction hypothesis [10] must possess simplex sets of nonrefinable effects and states to be ontologically noncontextual as defined by Spekkens [3]. We then discuss the scenario in which the no-restriction hypothesis is relaxed and show that this assumption is pivotal to the nonclassicality of GPTs describing single systems.

The practical significance of our analysis becomes clear via a markedly interesting example. 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 simulable [15–18]. This possibility is removed by providing only a single suitable nonstabilizer input state (or measurement) that “magically” enables fault-tolerant universal quantum computing [18–22]. 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 concerns the first phenomenon which is a prerequisite of the second.

The study of resources for information processing purposes commonly begins with assuming an underlying theory. Thinking of quantum theory, this is beautifully done within the formalism of quantum resource theories [23] such as entanglement [24], athermality [25–30], coherence [31–34], asymmetry [35–37], non-Markovianity [38], and dynamical correlations [39]. Here, we instead relax the assumption of a specific underlying theory by adopting the generic formalism of GPTs and investigate how the presence of a single resourceful operational element may conflict with the structure of noncontextual GPTs leading to potential nonclassical information processing advantages [5]. We use our first result to establish the necessary and sufficient criterion to distinguish between classicality and nonclassicality of a single resourceful measurement or state. Moreover, as a corollary, we show that in any ontologically noncontextual GPT complying with a Gleason-type theorem and the no-restriction hypothesis, modulo the possible coarse-grainings, there can only exist one measurement for any fixed number of outcomes.
II. PROLOGUE

A. A Gleason-type Theorem for GPTs

Through the rest of this paper, we are only interested in the prepare-and-measure experiments on single systems that can be described in finite dimensions by a finite number of measurement outcomes. Operationally, the primitive elements of our physical description in any such experiment are the laboratory prescriptions for preparations and measurements of the system of interest forming the collections \( \mathcal{P} := \{ P_k \} \) and \( \mathcal{M} := \{ M_j \} \), respectively [3]. The fundamental goal in theoretical physics is to establish assignments between these elements and mathematical objects endowed with a set of self-consistent rules that tell the experimenter how to determine the outcome probabilities in each measurement. A first step that is generic to any such approach is that the measurement processes with the same name of outcomes must be described by a measurable space \( (\Omega, \omega) \), where \( \Omega \) is the finite set of all outcomes and \( \omega \) is the \( \sigma \)-algebra of events on it.

Let us show the rest of this procedure within the framework of GPTs [8–14]. A GPT is constructed from a vector space \( \mathcal{V} \) for which one assumes only the primary properties of being ordered and real to hold, and which is endowed with an inner-product represented by \( \langle \cdot, \cdot \rangle \). In quantum theory, this vector space is the Banach space \( \mathcal{L}(\mathcal{H}) \) of all bounded linear Hermitian operators on a Hilbert space \( \mathcal{H} \).

The set of allowed states in a GPT satisfying the no-restriction hypothesis [10] is physically valid, meaning that they correspond to realizable states of the system. First, the probability rule of the GPTs is traditionally assumed a priori to be given by an inner product. Here, however, we chose a different route by deriving it from the properties of the effect spaces, following in footsteps of Gleason [44], Busch [42], and Caves et al [43]. Second, it is argued

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 \( q(A) = (A, B) \) for all \( A \in \mathcal{V} \), for a unique \( B \in \mathcal{V} \) which is normalized in the sense that \( (U, B) = 1 \).

Proof. Following the approach of Refs. [42, 43], 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 \( (E + \alpha E) \) and \( (E' + \beta E) \) imply \( q(E') = q(E') + q(\alpha E) \) and \( q(E') = 0 \) 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, \((\alpha_i)\) and \((\beta_i)\), respectively, where both converge to the same irrational value \( \gamma \in [0, 1] \). From order preserving property of \( q \) and its homogeneity over rational numbers we obtain \( \alpha_i q(E) = q(\alpha_i E) \leq q(\gamma E) \leq \beta_i q(\gamma) \). Then by the pinching theorem we have \( q(\gamma E) = q(\gamma) \), 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 \( q(U) = \langle U, B \rangle = 1 \).

Definition 1. The set of allowed states in a GPT satisfying the no-restriction hypothesis is given by

\[ \mathcal{S} := \{ \varrho \in \mathcal{V} | \langle E(X), \varrho \rangle \geq 0 \forall E(X) \in \mathcal{E}, (U, \varrho) = 1 \} \tag{1} \]

The GPT’s state space thus belongs to the dual cone \( \mathcal{E}^* \) to the cone of measurements and it is also pointed (i.e. \( \mathcal{E} \cap \mathcal{E}^* = \{ 0 \} \)) and generating (i.e., span \( \mathcal{E}^* = \mathcal{V} \) [9, 10, 12, 14].

There are two important aspects of our GPT construction different from what has previously been considered within literature. First, the probability rule of the GPTs is traditionally assumed a priori to be given by an inner product. Here, we chose a different route by deriving it from the properties of the effect spaces, following in footsteps of Gleason [44], Busch [42], and Caves et al [43]. Second, it is argued...
that the no-restriction hypothesis is of no physical basis and thus, it is desirable to drop it from GPT constructions. This approach has been taken, for example, in Refs. [11, 45] and extensively discussed by Janotta and Lal in Ref. [46]. Note that, such a “relaxation” comes at a price: one has to assume a priori also the assignment of preparations to the elements of the vector space underlying the theory, meaning trading one assumption for another. We have instead assumed the no-restriction hypothesis as a result of which the state space is determined by the effect space using the probability rule. Indeed, there is no objection to not imposing the no-restriction hypothesis, but we think the latter is more of a restriction rather than a relaxation as justified by the following proposition.

**Proposition 1.** Any GPT \( \mathcal{T}_{\text{sub}} \) defined on a vector space \( \mathcal{V} \) and identified by the pair \( \mathcal{T}_{\text{sub}} := (\mathcal{E}_{\text{sub}}, \mathcal{F}_{\text{sub}}) \) 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 \( \mathcal{T}_{\text{sub}} \subseteq \mathcal{T} \), where \( \mathcal{T} \) is defined as in Eq. (1), i.e.,

\[
\mathcal{T} := \{ \rho \in \mathcal{V} | \langle \rho, E(X) \rangle \geq 0 \forall E(X) \in \mathcal{E}_{\text{sub}}, \langle U, \rho \rangle = 1 \}.
\]

Hence, \( \mathcal{T}_{\text{sub}} \) is a subtheory of \( \mathcal{T} := (\mathcal{E}, \mathcal{F}) \), where \( \mathcal{T} \) satisfies the no-restriction hypothesis; see Fig. 1a. Alternatively, one can fix the state space and define the set of effects as

\[
\mathcal{E} := \{ E \in \mathcal{V} | \langle \rho, E(X) \rangle \in [0,1] \forall \rho \in \mathcal{E}_{\text{sub}} \}.
\]

Clearly this time \( \mathcal{E}_{\text{sub}} \subseteq \mathcal{E} \) and thus \( \mathcal{T}_{\text{sub}} \) is a subtheory of \( \mathcal{T}' := (\mathcal{E}, \mathcal{F}_{\text{sub}}) \), where \( \mathcal{T}' \) satisfies the no-restriction hypothesis; see Fig. 1b.

Finally, any GPT \( \mathcal{T}'' := (\mathcal{E}'', \mathcal{F}'') \) such that \( \mathcal{E}_{\text{sub}} \subseteq \mathcal{E} \subseteq \mathcal{E}'', \mathcal{F}_{\text{sub}} \subseteq \mathcal{F} \subseteq \mathcal{F}'' \) and \( \mathcal{E}' \) and \( \mathcal{E}'' \) are dual sets, contains \( \mathcal{T}_{\text{sub}} \) as its subtheory; see Fig. 1c.

As a result, we call any GPT that does not satisfy the no-restriction hypothesis simply a subGPT. Any subGPT can thus be viewed as many extended GPTs that satisfy the no-restriction hypothesis and on which appropriate additional restrictions are imposed. To give an example, consider quantum mechanics as a specific GPT that satisfies the no-restriction hypothesis and from which the Gaussian quantum mechanics is obtained by restricting the effect and state spaces to those elements that possess Gaussian Wigner representations [47]. Gaussian quantum mechanics is thus, by construction, an example of a GPT that does not satisfy the no-restriction hypothesis, i.e. it is a subGPT. It is also clear that quantum theory is not the unique GPT containing Gaussian quantum mechanics. In particular, classical statistical mechanics (or Liouville mechanics) also contains Gaussian quantum mechanics as its subtheory [47].

Within the rest of this paper, unless otherwise stated, we assume that GPTs comply with the no-restriction hypothesis.

**B. Ontological Models**

Recall that the primitive objects in the operational description are preparations and measurements and GPTs’ minimal aim is to provide a recipe for mathematically representing and composing these objects to predict the measurement outcomes at least probabilistically. In contrast, there exists a second approach known as ontological models whose purpose is to describe physical phenomena using elements of reality. An ontological model assumes that there exists an underlying ontic variable space \( \Upsilon \). The operational elements, i.e. preparations and measurements, generically correspond to probabilistic preparations and measurements of the ontic variable and thus, they are represented by probability distributions and indicator functions on \( \Upsilon \), respectively. More precisely, \( \Upsilon \) together with the \( \sigma \)-algebra \( \nu \) on it form a measurable space. Given the spaces \( \mathcal{Y} \) and \( \mathcal{D} \) of all probability measures on \( (\Upsilon, \nu) \) and \( (\Omega, \omega) \), respectively, the ontological model hypothesizes that there exist maps \( \mu : \mathcal{P} \rightarrow \mathcal{Y} \) and \( \xi : \mathcal{M} \rightarrow \mathcal{D} \) that assign the ontic state \( \mu_p \) to the preparation procedure \( P \) and the ontic measurement \( \xi_M \) to the measurement procedure \( M \). Similarly to PVVMs, it can be justified that the ontic maps \( \mu \) and \( \xi \) must be convex linear [3]. Each ontic state \( \mu_p \) is in fact a probability measure on \( \nu \) so that

\[
\mu_p : \nu \rightarrow [0,1] \quad \text{and} \quad \int_{\Upsilon} d\mu_p(\lambda) = 1. \quad (2)
\]

Each ontic measurement \( \xi_M \) on the other hand is a probability measure on \( \omega \) for every \( \lambda \in \Upsilon \) while it is a measurable function on \( \Upsilon \) for every \( X \in \omega \) so that

\[
\xi_M : \omega \times \Upsilon \rightarrow [0,1] \quad \text{and} \quad \xi_M(\Omega|\lambda) = 1 \quad \forall \lambda \in \Upsilon. \quad (3)
\]

Then, the probability of a particular event \( X \) in a measurement \( M \) given the preparation \( P \) can be obtained via Bayes’ rule,

\[
p(X|P, M) = \int_{\Upsilon} d\mu_p(\lambda)\xi_M(X|\lambda). \quad (4)
\]
FIG. 2. The diagram of the construction of different mathematical frameworks for operational descriptions of experiments. The dotted arrow represents the detour approach for building NCOMs by constructing ontological models for GPTs, as described within the main text. A GPT that allows for the dots to be joined and a NCOM to be constructed is called ontologically noncontextual.

C. Contextuality

Either choosing a GPT or an ontological model to describe an experiment (see Fig. 2), noncontextuality (as defined below) is a desirable property of the description. First, let us define statistical equivalence for preparation and measurement procedures as a plausible assumption [3, 40]. Two preparations $P_1, P_2 \in \mathcal{P}$ are statistically indiscernible and thus equivalent, $P_1 \equiv P_2$, 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 hence equivalent, $M_1 \equiv M_2$, 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 definitions naturally partition the collections of preparations and measurements into equivalence classes $e(P)$ and $e(M)$ for each preparation $P$ and each measurement $M$. The particular way in which a state or measurement is experimentally realized thus corresponds to an element within an equivalence class and it is called a context. The broad assumption of noncontextuality of a statistical model for experiments is that our models for the physical phenomena aiming only at reproducing the statistics should depend only on the contexts’ equivalence classes rather than individual contexts themselves, because the statistics do not carry any information about the latter. The notion of noncontextuality in this broad sense can be found in some former literature, for example in Ref. [43].

Noncontextuality of GPTs in a broad sense thus means that the assignments of states and measurements to preparation and measurement procedures are $e(P) \mapsto \varrho \in \mathcal{X}$ for all $P \in \mathcal{P}$ and $e(M) \mapsto \{E(X)\} \subset \mathcal{E}$ for all $M \in \mathcal{M}$, that is

$$
P_1 \equiv P_2 \iff \varrho_1 = \varrho_2, 
M_1 \equiv M_2 \iff \{E(X)\} = \{E(X)\}.
$$

Remark 1. Noncontextuality in the broad sense is built-in to our GPT construction via the definition of PVVMs and Theorem 1; see Fig 2.

Similarly, an ontological model which is noncontextual in the broad sense is called a noncontextual ontological model (NCOM) and satisfies $e(P) \mapsto \mu \in \mathcal{Y}$ for all $P \in \mathcal{P}$ and $e(M) \mapsto \xi \in \mathcal{D}$ for all $M \in \mathcal{M}$, that is

$$
P_1 \equiv P_2 \iff \mu_1 = \mu_2, 
M_1 \equiv M_2 \iff \{\xi(M)(X|\lambda)\} = \{\xi(M)(X|\lambda)\}.
$$

Remark 2. Every NCOM trivially is a GPT the underlying vector space of which is the space of real functions on $\mathcal{Y}$. However, the converse is not true, that is, not every GPT can be interpreted as a NCOM. Our aim is to determine which GPTs do and which ones do not admit a NCOM; see also Fig 2.

All classical theories are GPTs that can minimally be regarded as noncontextual probabilistic theories, hence NCOMs. Therefore, it is plausible to define experimental scenarios to be “classical” if and only if there exists a NCOM that fully simulates the statistics obtained in all allowed preparations and measurements [5].

III. MAIN RESULTS

A. Ontological (Non)Contextuality of GPTs

As we have inferred from experiments to date, that to take a direct route and build a NCOM (i.e., a classical theory) 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 2, and build ontological models of GPTs describing the world, noting that such models will necessarily inherit the noncontextuality from the theory leading to NCOMs. For the explicit example of quantum theory as a GPT, our question translates into whether or not quantum theory admits a NCOM, the answer to which is in negative [4, 48]. To take such a bypath in the case of a generic GPT, we simply replace the preparation and measurement procedures $P$ and $M$ in the construction of ontological models described earlier with their representatives within the given theory, $\varrho \in \mathcal{X}$ and $\{E(X)\} \subset \mathcal{E}$, respectively. Hence, for an ontological model of a GPT, there should exist injective maps $\eta: \mathcal{X} \to \mathcal{Y}$ and $\zeta: \mathcal{E} \to \mathcal{D}$ that assign the unique ontic state $\eta_0$ to each state

Note that, if the maps from the GPT to the ontological model are not injective then, there will exist multiple states (effects) that are mapped to the same ontic state (indicator function). Since each GPT state (effect) corresponds to an equivalent class of operational preparations (measurement events), it follows, in turn, that some of such equivalence classes will be mapped to the same ontic element. The latter cannot be true unless the ontological model has an intrinsic coarse-graining of preparations or measurements, meaning that the resulting ontological model is not maximally informative.

Furthermore, these maps must be invertible on their codomains. Suppose that this is not the case, which can happen only if there exist nonunique ontic states (indicator functions) in the codomain of $\eta (\zeta)$ that correspond to the same GPT state (effect). Now again, since each GPT state (effect) corresponds to an equivalent class of operational preparations (measurement events), it follows that there will exist equivalence classes that are represented by nonunique ontic elements. The latter cannot be true unless the ontological model is contextual, which is a contradiction.
vector \( \varrho \) and the unique ontic measurement \( \zeta_E \) to each PVVM \( E \) such that for all \( \lambda \in \Upsilon \) and all events \( X \in \omega \),

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

and satisfy

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

The probability of a particular event \( X \) in a measurement \( M \) given the preparation \( P \) should then be obtained as

\[
p(X|P,M) = p(X|\varrho,E) = \langle \varrho, E(X) \rangle = \int_\Upsilon d\eta_\varrho(\lambda) \zeta_E(X|\lambda).
\]

Evidently, given a preparation procedure \( P \in \mathcal{P} \) the ontological model defined above describes all the elements of its corresponding equivalence class \( e(P) \) with a single ontic state \( \mu_P(\lambda) = \mu_{e(P)}(\lambda) = \eta_\varrho(\lambda) \). A similar statement holds for measurements as well.

We remark that the focus of the present paper is the detour approach commonly considered within recent literature, in which a noncontextual theory usually refers to an operational theory that admits a NCOM [3]. For the sake of clarity, here we call such a theory ontologically noncontextual and reserve the term noncontextual theory (or model) for the older notion.

**Remark 3.** Any ontologically noncontextual theory is noncontextual, however, the converse is not true. For instance, quantum theory is noncontextual in the broad sense while it is not ontologically noncontextual [4, 48].

In the following, we establish the necessary and sufficient condition for a GPT to be ontologically noncontextual.

1. **GPTs that satisfy the no-restriction hypothesis**

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 \( \mathcal{Y} \), \( \eta \) and \( \zeta \) can uniquely be extended to the whole space \( \mathcal{Y} \). Then, using Riesz’s theorem we find that they must be of the forms

\[
\eta_\varrho(\lambda) = \langle \varrho, F(\lambda) \rangle \quad \text{and} \quad \zeta_E(X|\lambda) = \langle E(X), D(\lambda) \rangle,
\]

for \( F(\lambda), D(\lambda) \in \mathcal{Y} \). Then, in view of Eq. (10), satisfying Eq. (8) requires that \( \int_\Upsilon dF(\lambda) = U \) and \( \int_\Upsilon D(\lambda) = 1 \) for all \( \lambda \in \Upsilon \). Thus, \( \mathcal{F} := \{F(\lambda)\} \) resembles a PVVM whereas \( \mathcal{D} := \{D(\lambda)\} \) is akin to a subset of the GPT’s state space. The probability rule of Eq. (9) then implies that the pair \( (\mathcal{F}, \mathcal{D}) \) forms a frame-dual-frame [48, 49] for the vectors in \( \mathcal{Y} \) such that

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

After taking into account that, by the no-restriction hypothesis, state vectors are dual to measurement vectors and a few more simple steps we arrive at the pivotal result of the present paper.

**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 \((\mathcal{E}, \mathcal{F})\). That is, \( \mathcal{E} = \text{conv} \mathcal{F} \) and \( \mathcal{F} = \text{conv} \mathcal{D} \).

**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. (1) and that \( \langle U, D(\lambda) \rangle = 1 \) for all \( \lambda \in \Upsilon \), we find that \( D(\lambda) \in \mathcal{F} \), which implies \( \text{conv} \mathcal{F} \subseteq \mathcal{S} \). Second, by imposing the second requirement of a noncontextual ontological model on Eq. (11), that is, \( \eta_\varrho(\lambda) \geq 0 \) for all \( \varrho \in \mathcal{E} \) and all \( \lambda \in \Upsilon \), we see that \( \mathcal{S} \subseteq \text{conv} D(\lambda) \). Combining together, it must hold true that \( \text{conv} \mathcal{E} = \mathcal{F} \) and thus, the set \( \mathcal{F} \) 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 \( \varrho \in \mathcal{E} \) 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 \( F(\lambda) \in \mathcal{S} \) and thus, \( \text{conv} \mathcal{S} \subseteq \mathcal{E} \). Next, using the fact that \( \zeta_E(X|\lambda) \geq 0 \) for all \( E \in \mathcal{E} \) and all \( \lambda \in \Upsilon \) in Eq. (11) we find that \( \mathcal{E} \subseteq \text{conv} \mathcal{F} \). Together, we have \( \mathcal{E} = \text{conv} \mathcal{F} \), meaning that the set \( \mathcal{F} \) is a generating set of the GPT’s set of allowed effects \( \mathcal{E} \).

In light of Eq. (11), it is straightforward to conclude from Lemma 1 that the extreme states \( \varrho \in \mathcal{E} \) and the extreme effects \( E(X) \in \mathcal{F} \) must be represented by Dirac measures over the ontic space \( \Upsilon \), that is, (i) \( \exists \lambda_0 \in \Upsilon \) for some \( \lambda_0 \), and (ii) \( \mathcal{F} \ni E(X) \overset{\zeta}{\longrightarrow} \delta_{\lambda_0 E(X)}(\lambda) \) for some \( \lambda_0 \in \Upsilon \), where \( \delta_{\lambda_0}(\beta) \) equals 0 if \( a \notin \beta \) and equals 1 if \( a \in \beta \) for any measurable subset \( \beta \). The condition (i) refers to a property called ontic determinism, which means a pure preparation (i.e. those that cannot be refined into other preparations [10]) represented by a pure state in a GPT determines the ontic state in the corresponding NCOM to infinite precision. Condition (ii), on the other hand, is referred to as outcome determinability, that is, in sharp measurements [50] 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 a unique decompositions into nonrefinable extreme elements. Specifically, it is clear from Eq. (11) that requiring a unique ontic representation \( \eta_\varrho(\lambda) \) for a state vector \( \varrho \) as mentioned in Sec. III A simply means a unique convex decomposition into pure states. Extra care needs to be taken for making a similar argument for effects, because the extreme points of the effect space can themselves be (nonconvexly) decomposed into other extreme points due to the possibility of coarse-graining. For example, the unit ontic effect 1 can be decomposed into disjoint ontic effects in a nonunique manner even within a NCOM. It is, however, sufficient to consider only the sharp effects that cannot be refined (or atomic effects [10]) which are counterpart to pure state vectors. Denoting the set of all such effects by \( \mathcal{E}_{nr} \subseteq \mathcal{F} \), it follows that the convex decomposition of any effect into elements of \( \mathcal{E}_{nr} \) must be unique. The following theorem characterizes GPTs for which these considerations are met and thus, they are on-
Theorem 2. A GPT that is built upon the Gleason-type Theorem 1 and satisfies the no-restriction hypothesis is ontologically noncontextual if and only if its pure states and nonrefrable sharp effects each form a complete basis for the container space $\mathcal{V}$. Equivalently, the GPT must be simplicial meaning that $\mathcal{S}$ and $\text{conv}\mathcal{E}_{\text{nr}}$ are simplexes.

Proof. 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 $C \in \text{conv}\mathcal{A}$ such that its convex decomposition into elements of $\mathcal{A}$ is not unique.

Proof of Lemma 2. 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 $\mathcal{A}/A_j$. 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 $B \in \text{conv}\mathcal{A}$ such that $B \notin \text{conv}\mathcal{A}/A_j$ and

$$B = \sum_{i \neq j} \beta_i A_i, \quad \sum_i \beta_i = 1.$$  \hspace{1cm} (12)

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 $A^* = \sum_{i \neq j} \alpha_i A_i = (1 - p)A^*$, where $p = \sum_{i \neq j} \alpha_i A_i$. Then, for the expansion coefficients of $B(p)$ it holds true that $\sum_{i \neq j} \beta_i(p) = p \sum_{i \neq j} \alpha_i A_i + (1 - p)A^*$. By defining $\pi : = |\sum_{i \neq j} \alpha_i|$, we find

$$\overline{p} := p = \begin{cases} 0, & \text{if } \pi = 0, \\ 1/\sum_{i \neq j} \alpha_i, & \text{otherwise}, \end{cases}$$  \hspace{1cm} (13)

for which $\sum_{i \neq j} \beta_i(\overline{p}) = 1$. Evidently, $B(\overline{p}) \notin \text{conv}\mathcal{A}/A_j$ by construction. However, because there exists at least one $\beta_i(\overline{p}) = \alpha^* \beta_i < 0$, it also holds true that $B(\overline{p}) \notin \text{conv}\mathcal{A}/A_j$. We thus can set $B = B(\overline{p})$.

Given the operator $B$ with the properties as in Eq. (12), we consider two sets of indices: $\mathcal{J}^- := \{i \neq j | \beta_i < 0\}$ and $\mathcal{J}^+ := \{i \neq j | \beta_i > 0\}$ and define the operator

$$C := \frac{1}{N} (B + \sum_{i \in \mathcal{J}^-} |\beta_i| A_i) = \frac{1}{N} \sum_{i \in \mathcal{J}^+} \beta_i A_i,$$  \hspace{1cm} (14)

in which $N = \sum_{i \in \mathcal{J}^+} |\beta_i|$. We see that, both sides of Eq. (14) 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 non-unique decomposition in terms of $\mathcal{E}_{\text{nr}} \subset \mathcal{F}$. This means the effect $C$ possesses a non-unique 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 $\mathcal{F}$ form a nonconvexly overcomplete basis.

2. GPTs that fail to satisfy the no-restriction hypothesis

In Sec. II A, we discussed how any GPT that does not satisfy the no-restriction hypothesis can be thought of as a restricted subtheory of a GPT that complies with this assumption. Now, an interesting question to ask is what we can say about the ontological (non)contextuality of subGPTs. The following is rather easy to show.

Theorem 3. All subGPTs of an ontologically noncontextual GPT are ontologically noncontextual.

Proof. Since for any subGPT $\mathcal{S}_{\text{sub}} = (\mathcal{E}_{\text{sub}}, \mathcal{S}_{\text{sub}})$ 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 assumed ontologically noncontextual GPT, then for all elements $E(X) \in \mathcal{E}_{\text{sub}}$ and $\rho \in \mathcal{S}_{\text{sub}}$ one can easily use the same ontic assignments $\zeta(E(X) | \lambda)$ and $\eta(\lambda)$ of the full GPT $\mathcal{F} = (\mathcal{E}, \mathcal{S})$. These assignments trivially reproduce the statistics of the subGPT.

By putting together Remark 2 and Theorem 3 we arrive at the following.

Theorem 4. Any GPT that does not satisfy the no-restriction hypothesis admits a NCOM if and only if it is a subtheory of an ontologically noncontextual GPT.

Proof. The if direction follows from Theorem 3. The only if direction can be shown as follows. Assume that the subtheory $\mathcal{S}_{\text{sub}} = (\mathcal{E}_{\text{sub}}, \mathcal{S}_{\text{sub}})$ 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 $\mathcal{Y}$. We note that we can always assume that the effect and state spaces of the subGPT span its underlying vector space $\mathcal{V}$. As a result, $\eta_{\text{sub}}$ and $\zeta_{\text{sub}}$ can be uniquely extended to bijective maps $\eta$ and $\zeta$ over the whole $\mathcal{Y}$. We also know that, due to being probability measures on $\mathcal{Y}$, 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. 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 $\mathcal{E}$ and $\mathcal{S}$ are dual and satisfy the no-restriction hypothesis. Moreover, $\mathcal{F} = (\mathcal{E}, \mathcal{S})$ is ontologically noncontextual by construction. Finally, $\mathcal{S}_{\text{sub}}$ is a subtheory of $\mathcal{F}$. 

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 noncontectual GPT.

In a closely related and simultaneous work, Schmid et al [51] obtained a similar result to Theorem 4 wherein sub-GPTs admitting a NCOM are denoted as simplex-embeddable GPTs.

We discussed following Proposition 1 that removing the no-restriction hypothesis, in a sense, is more of a restriction rather than a relaxation. Another argument supporting the physical importance of the no-restriction hypothesis can be given as follows. The nonsimplicial subGPTs that do not satisfy the no-restriction hypothesis may or may not be ontologically noncontextual, because they may or may not be sub-GPTs of classical GPTs. However, we learn from Theorem 2 that once a nonsimplicial subGPT is enforced to satisfy the no-restriction hypothesis by enlarging its effect or state spaces in a fixed container vector space $\mathcal{V}$, it will definitely result in an ontologically contextual, hence nonclassical, extended GPT. As a result, although assuming the no-restriction hypothesis may be unnecessary, it suffices for the nonclassicality of nonsimplicial GPTs for single systems. A more challenging problem is to obtain the necessary condition disallowing simplicial extensions of subGPTs.

3. An Example: Spekkens’ Toy Theory

As an interesting example of the application of Theorems 2, 3, and 4, we analyse the noncontextuality of the Spekkens’ toy theory [52].

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 $(1, 0, 0, 0)^T$, $(0, 1, 0, 0)^T$, $(0, 0, 1, 0)^T$, and $(0, 0, 0, 1)^T$, respectively. After enforcing the knowledge balance principle on the theory, a subtheory will be obtained the pure states of which are given as “1 $\lor$ 2”, “1 $\lor$ 3”, “1 $\lor$ 4”, “2 $\lor$ 3”, “2 $\lor$ 4”, and “3 $\lor$ 4” [52]. Here, $\lor$ denotes the disjunction or “OR” operator. These states can be represented by column vectors $(1/2, 1/2, 0, 0)^T$, $(1/2, 0, 1/2, 0)^T$, $(1/2, 0, 0, 1/2)^T$, $(0, 1/2, 1/2, 0)^T$, $(0, 1/2, 0, 1/2)^T$, and $(0, 0, 1/2, 1/2)^T$, respectively, which form an octahedron that is circumscribed by the GPT’s state space; see Fig. 3. The set of reproducable and nonrefinable PVVMs also reduces to $\{1 \lor 2, 3 \lor 4\}$, $\{1 \lor 3, 2 \lor 4\}$, and $\{1 \lor 4, 2 \lor 3\}$, that can be represented by $\{(1, 1, 0, 0), (0, 0, 1, 1)\}$, $\{(1, 0, 1, 0), (0, 1, 0, 1)\}$.

2 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 [52]. However, this fact is irrelevant to our analysis.
and \( \{ (1, 0, 0, 1), (0, 1, 1, 0) \} \), 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. Nevertheless, as a subtheory of an ontologically noncontextual theory, Spekkens’ subGPT is ontologically noncontextual by Theorem 3.

Is it possible to consider Spekkens’ toy model as a subtheory of an ontologically contextual GPT? The answer is in affirmative. Referring to the discussion of Sec. III A 2, it is sufficient to enlarge the subtheory to a GPT that satisfies the no-restriction hypothesis, say by fixing the effect space of the toy model to \( \mathcal{E}_{\text{Spekkens}} \) and assuming the no-restriction hypothesis to define the state space of the GPT as the dual set to \( \mathcal{E}_{\text{Spekkens}} \) according to Eq. (1). This will, for instance, account for vectors \((1, 1, 1, 1)^T/2, (1, -1, 1, 1)^T/2, (1, 1, -1, 1)^T/2, \) and \((1, 1, 1, -1)^T/2 \) as valid pure states on top of the extreme points of the ontological simplex. \((0, 0, 0, 0)^T, (0, 1, 0, 0)^T, (0, 0, 1, 0)^T, \) and \((0, 0, 0, 1)^T \). It follows that the resulting state space is not a simplex in \( \mathbb{R}^4 \) and thus, by Theorem 2, the resulting GPT is ontologically contextual. We emphasize that the extended ontologically contextual theory obtained in this way need not satisfy the knowledge balance principle, and indeed, it does not. However, by construction, Spekkens’ toy model is a subGPT of this extended ontologically contextual theory.

B. Resources and Contextuality in GPTs

Having the scene set, we are ready to consider the concept of resources in GPTs. Imagine experiments with the sets of preparations \( \mathcal{P} \) and measurements \( \mathcal{M} \), and an experimenter who has devised a GPT which is tailored to explain the statistics obtained in her experiments. Identifying this GPT with the pair \( \mathcal{T} := (\mathcal{E}, \mathcal{S}) \) of operationally allowed PVVMs and states, we call the set of all possible measurements \( \mathcal{E} \) and the set of free measurements and the set of all possible preparations \( \mathcal{S} \) represents the free states.

We may remark here on the difference between our definition of free elements with the one common to quantum information literature [23–39]. In quantum resource theories free states and measurements are subsets of the sets of all possible states and measurements that are defined with respect to a set of operationally sufficiently low-cost-to-perform operations, e.g. LOCC in entanglement theory. Here, however, all possible preparations and measurements are defined to be free. Indeed all possible operations are also assumed to be free, though we are not interested in operations here. The nonfree (or resourceful) states and measurements in our scenario are those yet to be discovered in nature and thus not specified in the GPT adopted by the experimenter.

Now, suppose that the experimenter comes up with another measurement procedure that was not previously known to be possible, say \( M^* \in \mathcal{E}(M) \) for all \( M \in \mathcal{M} \). This can happen, for example, upon the invention of a special measuring equipment. We call such a bonus element a resourceful measurement. Naturally, the experimenter has to come up with a new extended theory \( \mathcal{T}^* := (\mathcal{E}^*, \mathcal{S}^*) \) to reproduce also the new measurement data obtained in \( M^* \). We ask, “how does \( \mathcal{T}^* \) compare to \( \mathcal{T} \) in terms of (non)classicality?”

By Theorem 2, given the whole packages of new measurements \( \mathcal{M}^* \) and new preparations \( \mathcal{P}^* \), 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 meaning that there exists a classical model capable of fully reproducing all the observed statistics. In such scenarios, even though the measurement \( M^* \) is a resource, all the measured statistics can be fully explained in classical terms. Hence, we call such bonus measurements classical resources. We thus further narrow down our previous question to “what are the necessary and sufficient conditions for a resource to enforce a nonclassical extension of \( \mathcal{T} \)?” When such a resource measurement exists, we call it a nonclassical resourceful measurement. We also refer the interested reader to Kunjwal and Spekkens [6] and Schmid et al [7] for an approach to detecting such nonclassical statistics using noncontextuality inequalities. The answer to this question is presented in the following theorem with an intuitive geometrical interpretation as shown in Fig 4.

Theorem 5. 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 (i.e. ontologically noncontextual) GPT \( \mathcal{T} = (\mathcal{E}, \mathcal{S}) \). Given a single nonrefinable bonus measurement \( M^* \notin \mathcal{M} \) and assuming that the resulting extended theory \( \mathcal{T}^* := (\mathcal{E}^*, \mathcal{S}^*) \) must satisfy the no-restriction hypothesis, the followings are equivalent:

1. \( \mathcal{T}^* \) is ontologically contextual;
2. \( M^* \) is a nonclassical resourceful measurement;
3. The PVVM \( E^* \) nonconvexly overcomes the nonrefinable effects \( \mathcal{E}_{\text{nr}} \) into \( \mathcal{E}_{\text{nr}}^* \);
4. \( E^* \) lies within \( \mathcal{V} \) but \( E^* \notin \mathcal{E} \).

Proof. That 1 if and only if 2 follows from the definition of nonclassical resources given above. That 1 if and only if 3 follows from Theorem 2 and that \( \mathcal{E}_{\text{nr}} \) is a complete basis for \( \mathcal{V} \). That 3 if and only if 4 follows from the assumption that \( \mathcal{E}_{\text{nr}} \) form a complete basis for \( \mathcal{V} \), hence adding an extra nonrefinable extreme point overcomes it. 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.

We can similarly analyze the case wherein the experimenter discovers a new preparation procedure \( P^* \) for a system. We also call such bonus preparation a resourceful preparation. Moreover, whenever \( P^* \) enforces an ontologically contextual extension of a classical theory, we call it a nonclassical resourceful preparation.

Theorem 6. Suppose that the set of free preparations \( \mathcal{P} \) is represented by the set of all states \( \mathcal{S} \) in some classical GPT \( \mathcal{T} = (\mathcal{E}, \mathcal{S}) \). Given a single bonus pure preparation \( P^* \notin \mathcal{P} \) and assuming that the resulting extended theory \( \mathcal{T}^* := (\mathcal{E}^*, \mathcal{S}^*) \) must satisfy the no-restriction hypothesis, the followings are equivalent:
and showed that any GPT built upon a Gleason-type theorem satisfying the no-restriction hypothesis is ontologically noncontextual if and only if it is simplicial. We also discussed extensively the case of subGPTs that do not comply with the no-restriction hypothesis. Our results shows 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 to each possibility already exist, e.g. dropping the no-restriction hypothesis by imposing epistemic restrictions to a given GPT that results in subGPTs, as in Gaussian subtheory of quantum mechanics [47], or giving up on the ontological noncontextuality as in full quantum theory, or capitulating incompatible measurements as in classical mechanics.

A secondary aspect of our work is the characterization of the power of individual resourceful operational elements in inducing ontologically contextual GPTs. A phenomena that has puzzled quantum information scientists for long is that the advantages obtained in some nonclassical information processing protocols, in particular, quantum computations, are commonly obtained from having access to merely one resourceful element.

Currently, the main approach to address the open problem of what is special about such single resource elements is to first assume that quantum theory is valid. Next, define a quantum resource theory framework [23], for example, entanglement [24] or coherence [31–34]. Then define a maximal resourceful state or measurement (e.g. the maximally entangled state or the maximally coherent state) with respect to some partial ordering of the elements within that resource theory as the representative resource, and then, explain the nonclassical advantage of the protocol under consideration using the resource property of that representative element. The usefulness of other resource elements are commonly tied with the possibility of transforming them into the maximal resourceful element using the free operations of the resource theory.

The above-mentioned approach, however, has some caveats. For instance, usually it is hard to find reasonable justifications as to why a particular resource theory is the relevant one, because various quantum resources may be assigned to a given element. For instance, any entangled state also contains quantum coherence. Moreover, the ordering of the resources is commonly partial, hence, the resource theory under consideration may fail to possess a unique representative maximally resourceful element [53]. Finally, in some extreme cases like multipartite entanglement conversion between most resources is almost impossible [54], leaving the process incomplete.

Here, in contrast, we have started a novel approach by directly accounting for the nonclassical power of resourceful elements by noting that many of nonclassical advantages, though possibly explainable by quantum formalism, do not depend on the specific underlying theory assumed. This is because, considering the particular example of quantum computation protocols, we can draw conclusions about their nonclassicality (similarly to the Bell scenario) by merely relying on the classical inputs and outputs to these protocols and the promises of the computational complexity theory. Taking into account this fact, we arrive at two conclusions. First, it fol-

---

**FIG. 4.** A heuristic representation of two hypothetical nonrefinable sharp PVVMs $E$ and $E^*$ as two-event measurements in a three dimensional space. The three nonrefinable elements \( \{ E(X_0), E(X_1), E^*(X_0) \} \) form a complete basis for this space so that the four \( \{ E(X_0), E(X_1), E^*(X_0), E^*(X_1) \} \) are overcomplete. An equivalent geometrical way to see the overcompleteness is that there are two different paths from zero to the unit element $U$ through nonrefinable elements; one by following black arrows passing through $E(X_0)$ that correspond to $E(X_0) + E(X_1)$ and one by following blue arrows passing through $E^*(X_1)$ that correspond to $E^*(X_1) + E^*(X_0)$. Clearly, there is no way of having two (or more) nonrefinable PVVMs without running into an overcomplete basis and an ontologically contextual GPT, as stated in Corollary 1. The magenta and green lines represent an idea of the dual cones containing the state spaces as delineated by Definition 1.

1. $\mathcal{F}^*$ is ontologically contextual;
2. $\mathcal{P}^*$ is a nonclassical resourceful preparation;
3. The state $\varrho^*$ nonconvexly overcompletes the states $\mathcal{F}$ into $\mathcal{F}^*$;
4. $\varrho^*$ lies within $\mathcal{V}$ but $\varrho^* \notin \mathcal{E}$.

**Proof.** The proof is similar to that of Theorem 5.

Theorems 5 and 6 thus identify the conditions for a single resource to dictate the use of a nonclassical model for an explanation of possible statistics in experiments. Theorem 5, however, also has a more striking consequence.

**Corollary 1.** In any ontologically noncontextual GPT that is built upon the Gleason-type Theorem 1 and satisfies the no-restriction hypothesis there exists only one sharp nonrefinable PVVM (on a fixed event space).

**Proof.** Suppose that $E = \{ E(X_i) \}$ and $E^* = \{ E^*(X_i) \}$ are two sharp nonrefinable PVVMs of an ontologically noncontextual GPT that is compliant with the no-restriction hypothesis. Then, by property (ii) of PVVMs we have that $E(\Omega) = U = \sum_i E(X_i) = \sum_i E^*(X_i)$ which means $E$ and $E^*$ are linearly dependent. Since $E(X_i)$ and $E^*(X_i)$ for all $X_i \in \omega$ are extreme points of $\mathcal{F}$ which span $\mathcal{V}$, it follows that either $E = E^*$ or the extreme points of $\mathcal{F}$ form an overcomplete basis for $\mathcal{V}$. The latter, however, contradicts with the GPT being ontologically noncontextual according to Theorem 2. Hence, the former must hold, i.e., $E = E^*$.

IV. DISCUSSION AND CONCLUSIONS

In conclusion we have considered the phenomenon of generalized contextuality in general probabilistic theories (GPTs)
lows that an advantage in a quantum scenario also implies an advantage in any postquantum theory, which can be well-formulated within the GPT framework. Second, we conclude that the resource formalism suitable for explaining the non-classical advantages should describe a fundamental theory-independent property, i.e., one that can be certified merely by relying on the classical measurement outcome statistics. Contextuality is one such a property that ipso facto forbids a classical description of the processes. Combining these two, in our opinion, the study of the 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 mentioned earlier.

ACKNOWLEDGMENTS

The author is thankful 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 Sér SAM Project at Swansea University, an initiated funded through the Welsh Government’s Sér Cymru II Program (European Regional Development Fund) and AQTION project (820495) funded by the European Union Quantum Technology Flagship.

[1] S. Kochen and E. P. Specker, The problem of hidden variables in quantum mechanics, 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 (2010).
[3] R. W. Spekkens, Contextuality for preparations, transformations, and unsharp measurements, Phys. Rev. A 71, 052108 (2005).
[4] R. W. Spekkens, Negativity and Contextuality are Equivalent Notions of Nonclassicality, Phys. Rev. Lett. 101, 020401 (2008).
[5] D. Schmid and R. W. Spekkens, Contextual Advantage for State Discrimination, Phys. Rev. X 8, 011015 (2018).
[6] R. Kunjwal and R. W. Spekkens, From the Kochen-Specker Theorem to Noncontextuality Inequalities without Assuming Determinism, Phys. Rev. Lett. 115, 110403 (2015).
[7] D. Schmid, R. W. Spekkens, and E. Wolfe, All the noncontextuality inequalities for arbitrary prepare-and-measure experiments with respect to any fixed set of operational equivalences, Phys. Rev. A 97, 062103 (2018).
[8] L. Hardy, Quantum Theory From Five Reasonable Axioms, (2001), arXiv:0101012 [quant-ph].
[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, (2011), arXiv:1104.2066.
[13] L. Masanes and M. P. Müller, A derivation of quantum theory from physical requirements, New Journal of Physics 13, 065001 (2011).
[14] P. Janotta and H. Hinrichsen, Generalized probability theories: what determines the structure of quantum theory? Journal of Physics A: Mathematical and Theoretical 47, 323001 (2014).
[15] A. Mari and J. Eisert, Positive Wigner Functions Rendext Classical Simulation of Quantum Computation Efficient, Phys. Rev. Lett. 109, 230503 (2012).
[16] 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).
[17] V. Veitch, S. A. Hamed Mousavian, D. Gottesman, and J. Emerson, The resource theory of stabilizer quantum computation, New J. Phys. 16, 013009 (2014).
[18] R. Raussendorf, J. Bermejo-Vega, E. Tuyhurst, C. Okay, and M. Zurel, Phase space simulation method for quantum computation with magic states on qubits, arXiv:1905.05374 (2019).
[19] D. Gottesman and J. L. Chuang, Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations, Nature 402, 390 (1999).
[20] S. Bravyi and A. Kitaev, Universal quantum computation with ideal Clifford gates and noisy ancillas, Phys. Rev. A 71, 022316 (2005).
[21] E. Knill, Quantum computing with realistically noisy devices, Nature 434, 39 (2005).
[22] M. Howard, J. Wallman, V. Veitch, and J. Emerson, Contextuality supplies the ‘magic’ for quantum computation, Nature 510, 351 (2014).
[23] E. Chitambar and G. Gour, Quantum resource theories, Rev. Mod. Phys. 91, 025001 (2019), 1806.06107.
[24] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Quantum entanglement, Rev. Mod. Phys. 81, 865 (2009).
[25] 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).
[26] M. Horodecki and J. Oppenheim, Fundamental limitations for quantum and nanoscale thermodynamics, Nature Communications 4, 2059 (2013).
[27] F. Brandão, M. Horodecki, N. Ng, J. Oppenheim, and S. Wehner, The second laws of quantum thermodynamics, Proceedings of the National Academy of Sciences 112, 3275 (2015).
[28] P. Faist, F. Dupuis, J. Oppenheim, and R. Renner, The minimal work cost of information processing, Nature Communications 6, 7669 (2015).
[29] M. Lostaglio, K. Korzekwa, D. Jennings, and T. Rudolph, Quantum coherence, time-translation symmetry, and thermodynamics, Phys. Rev. X 5, 021001 (2015).
[30] G. Gour, M. P. Müller, V. Narasimhachar, R. W. Spekkens, and N. Y. Halpern, The resource theory of informational nonequilibrium in thermodynamics, Physics Reports 583, 1 (2015).
[31] J. Aberg, Quantifying superposition, arXiv:0612146 (2006).
[32] T. Baumgratz, M. Cramer, and M. B. Plenio, Quantifying coherence, Phys. Rev. Lett. 113, 140401 (2014).
[33] A. Winter and D. Yang, Operational resource theory of coher-
ence, Phys. Rev. Lett. 116, 120404 (2016).
[34] A. Streltsov, G. Adesso, and M. B. Plenio, Colloquium: Quantum coherence as a resource, Rev. Mod. Phys. 89, 041003 (2017).
[35] G. Gour and R. W. Spekkens, The resource theory of quantum reference frames: manipulations and monotones, New Journal of Physics 10, 033023 (2008).
[36] G. Gour, I. Marvian, and R. W. Spekkens, Measuring the quality of a quantum reference frame: The relative entropy of frame-ness, Phys. Rev. A 80, 012307 (2009).
[37] 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).
[38] A. Rivas, S. F. Huelga, and M. B. Plenio, Quantum non-markovianity: characterization, quantification and detection, Reports on Progress in Physics 77, 094001 (2014).
[39] A. Rivas and M. Müller, Quantifying spatial correlations of general quantum dynamics, New Journal of Physics 17, 062001 (2015).
[40] A. Holevo, Probabilistic and Statistical Aspects of Quantum Theory (Edizioni della Normale, Pisa, 2011).
[41] P. Busch, M. Grabowski, and P. J. Lahti, Operational Quantum Physics (Springer Berlin Heidelberg, Berlin, Heidelberg, 1995).
[42] P. Busch, Quantum states and generalized observables: A simple proof of Gleason’s theorem, Phys. Rev. Lett. 91, 120403 (2003).
[43] C. M. Caves, C. A. Fuchs, K. K. Manne, and J. M. Renes, Gleason-type derivations of the quantum probability rule for generalized measurements, Foundations of Physics 34, 193 (2004).
[44] A. M. Gleason, Measures on the closed subspaces of a Hilbert space, Journal of Mathematics and Mechanics 6, 885 (1957).
[45] H. Barnum, J. Barrett, L. O. Clark, M. Leifer, R. Spekkens, N. Stepnik, A. Wilce, and R. Wilke, Entropy and information causality in general probabilistic theories, New Journal of Physics 12, 033024 (2010).
[46] P. Janotta and R. Lal, Generalized probabilistic theories without the no-restriction hypothesis, Phys. Rev. A 87, 052131 (2013).
[47] 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).
[48] C. Ferrie and J. Emerson, Framed Hilbert space: hanging the quasi-probability pictures of quantum theory, New J. Phys. 11, 063040 (2009).
[49] 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).
[50] G. Chiribella and X. Yuan, Measurement sharpness cuts non-locality and contextuality in every physical theory, (2014), arXiv:1404.3348v2.
[51] D. Schmid, J. H. Selby, R. W. Spekkens, E. Wolfe, and R. Kunjwal, The Characterization of Noncontextuality in the Framework of Generalized Probabilistic Theories, To appear (2019).
[52] R. W. Spekkens, Evidence for the epistemic view of quantum states: A toy theory, Phys. Rev. A 78, 032110 (2007).
[53] J. I. de Vicente, C. Spee, and B. Kraus, Maximally entangled set of multipartite quantum states, Phys. Rev. Lett. 111, 110502 (2013).
[54] D. Sauerwein, N. R. Wallach, G. Gour, and B. Kraus, Transformations among pure multipartite entangled states via local operations are almost never possible, Phys. Rev. X 8, 031020 (2018).