Characterization of Noncontextuality in the Framework of Generalized Probabilistic Theories

David Schmid,[1,*] John H. Selby,[1,2,†] Elie Wolfe,[1] Ravi Kunjwal,[1,3] and Robert W. Spekkens[1]

[1] *Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo, Ontario N2L 2Y5 Canada*

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

[3] *Centre for Quantum Information and Communication, Ecole polytechnique de Bruxelles, CP 165, Université libre de Bruxelles, Brussels 1050, Belgium*

(Received 27 January 2020; revised 28 July 2020; accepted 11 January 2021; published 25 February 2021)

To make precise the sense in which the operational predictions of quantum theory conflict with a classical worldview, it is necessary to articulate a notion of classicality within an operational framework. A widely applicable notion of classicality of this sort is whether or not the predictions of a given operational theory can be explained by a generalized-noncontextual ontological model. We here explore what notion of classicality this implies for the generalized probabilistic theory (GPT) that arises from a given operational theory, focusing on prepare-measure scenarios. We first show that, when mapping an operational theory to a GPT by quotienting relative to operational equivalences, the constraint of explainability by a generalized-noncontextual ontological model is mapped to the constraint of explainability by an ontological model. We then show that, under the additional assumption that the ontic state space is of finite cardinality, this constraint on the GPT can be expressed as a geometric condition which we term simplex embeddability. Whereas the traditional notion of classicality for a GPT is that its state space be a simplex and its effect space be the dual of this simplex, simplex embeddability merely requires that its state space be *embeddable* in a simplex and its effect space in the dual of that simplex. We argue that simplex embeddability constitutes an intuitive and freestanding notion of classicality for GPTs. Our result also has applications to witnessing nonclassicality in prepare-measure experiments.

DOI: 10.1103/PRXQuantum.2.010331

I. INTRODUCTION

In what precise sense does quantum theory necessitate a departure from a classical worldview? Although this is one of the central questions in the foundations of quantum theory, there is no consensus on its answer. Arguably the two most stringent notions of nonclassicality proposed to date are: the failure to admit of a locally causal ontological model (Bell's theorem) [1,2] and the failure to admit of a generalized-noncontextual ontological model [3]. Both of these are operationally meaningful notions of nonclassicality, in the sense that one can determine in principle whether a given set of operational statistics admits of a classical explanation by their lights, regardless of its consistency with quantum theory [4]. This implies that any experimental evidence for such nonclassicality imposes a constraint on any physical theory that hopes to be empirically adequate, including any putative successor to quantum theory.

For prepare-measure experiments on a single system, the notion of local causality is not applicable, and so, of the two notions, only generalized noncontextuality is a candidate for an operationally meaningful notion of classicality for such experiments. Elsewhere [5] it has been argued that its operational meaningfulness and its larger scope of applicability make the notion of generalized noncontextuality the best notion of classicality available today. Furthermore, it can be shown to subsume the central ideas behind several other notions of classicality, such as the existence of a nonnegative quasiprobability representation [6,7], or of a locally causal model [1,2]. Additionally, the *failure* of generalized noncontextuality has been shown to be behind certain notions of nonclassicality, such as anomalous weak values [8,9] and advantages for information processing [10–19].

Within the framework of ontological models [3], assuming generalized noncontextuality can be understood as assuming a version of a methodological principle for

*dschmid@perimeterinstitute.ca

†john.h.selby@gmail.com

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

theory construction due to Leibniz: the ontological identity of empirical indiscernibles (see Ref. [20] and the appendix of Ref. [21]). Given that Einstein made significant use of this principle when he developed the theory of relativity [20], it is seen to have impressive credentials in physics and therefore is a natural constraint to impose on ontological models. From this perspective, the impossibility of finding generalized-noncontextual ontological models is best understood as a failing of the framework of ontological models itself, and hence as a type of nonclassicality.

An operational theory provides an account of the experimental procedures accessible in the lab and the operational statistics they yield. A generalized probabilistic theory (GPT) [22,23] is obtained from an operational theory by discarding information about experimental procedures that can be varied without affecting the operational statistics. An ontological model of an operational theory or of a GPT is an attempt to provide a realist underpinning to these, that is, a causal account of the statistics they predict. We here characterize what the existence of a generalized-noncontextual ontological model of an operational theory implies about the geometry of the GPT associated to the operational theory. Ultimately, we prove the following result.

Theorem 1. *For a prepare-measure experiment, the operational theory describing it admits of a generalized-noncontextual ontological model on an ontic state space of finite cardinality if and only if the GPT describing it is* simplex-embeddable.

Simplex embeddability, defined rigorously in Definition 1 below, stipulates that there is a linear map that embeds the state space in a simplex and another linear map that embeds the effect space in the dual of this simplex, such that the pair of maps together preserve inner products.

Hence, if one takes explainability by a generalized-noncontextual model as one's notion of classicality for an operational theory, then one must take simplex embeddability as one's notion of classicality for a GPT. This is in contrast with the prevailing idea (see, e.g., Ref. [23]) that a GPT should be deemed classical if and only if its state space is a simplex and its effect space is the dual thereof, a condition which we term *simpliciality*. In our conclusions, we discuss the significance of the difference between simplex embeddability and simpliciality as notions of classicality for a GPT.

We begin by introducing the requisite preliminary concepts.

II. OPERATIONAL THEORIES

An *operational theory* is a minimal type of theory that stipulates, for a given system, a set of preparation procedures and measurement procedures that can be implemented on that system, denoted Preps and Mmts, respectively. These are conceptualized as lists of lab instructions that one could implement on the given system. We here find it useful to consider *operational effects*, defined as the tuple consisting of a measurement and an outcome thereof. We obtain the set of all operational effects, denoted $\mathsf{Effects}$, by considering the set of all outcomes k for each measurement M in the set Mmts. A particular operational effect will be denoted $[k|M]$. An operational theory stipulates a probability rule that determines the probability of obtaining operational effect $[k|M]$ given preparation P, denoted $\Pr([k|M], P)$. This probability rule must be compatible with certain relations that hold between the procedures [24]. For instance, if P_1 is described as a procedure that convexly mixes P_2 and P_3, with the choice determined by a coin-flipping mechanism, then $\Pr([k|M], P_1)$ must be equal to the corresponding mixture of $\Pr([k|M], P_2)$ and $\Pr([k|M], P_3)$ for all $[k|M]$ [22]. For operational effects, analogous constraints from convexity hold, as do additional constraints due to coarse-graining relationships among operational effects. For example, if one operational effect $[k_1|M_1]$ is described as being the coarse-graining of two others, $[k_2|M_2]$ and $[k_3|M_3]$, then $\Pr([k_1|M_1], P)$ must be the sum of $\Pr([k_2|M_2], P)$ and $\Pr([k_3|M_3], P)$ for all P. As we comment below, these constraints on $\Pr(_,_)$ have important consequences for the generalized probabilistic theory associated to the operational theory.

In all, an operational theory of a prepare-measure experiment on a single system is a triple $T := (\mathsf{Preps}, \mathsf{Mmts}, \Pr(_,_))$ satisfying these constraints.

Finally, we define the notion of *operational equivalence* of procedures [3]. Preparation procedures P and P' are said to be operationally equivalent, denoted $P \simeq P'$, if they give rise to the same statistics for all physically possible operational effects, that is, if $\Pr([k|M], P) = \Pr([k|M], P')$ for all $[k|M] \in \mathsf{Effects}$. Operational effects $[k|M]$ and $[k'|M']$ are said to be operationally equivalent, denoted $[k|M] \simeq [k'|M']$, if they give rise to the same statistics for all physically possible preparations, that is, if $\Pr([k|M], P) = \Pr([k'|M'], P)$ for all $P \in \mathsf{Preps}$.

III. GENERALIZED PROBABILISTIC THEORIES

The framework of generalized probabilistic theories provides a means of describing the landscape of possible theories of the world, as characterized (solely) by the operational statistics they predict [22,23]. Quantum and classical theories are included as special cases, but the framework also accommodates alternatives to these. Although the framework allows for sequential and parallel composition of processes, we focus on the fragment of a GPT that describes prepare-measure experiments on a single system.

A given GPT associates to a system a convex set of *states*, Ω. One can think of this set as being a generalization of the Bloch ball in quantum theory, where the states in the set are the normalized (potentially mixed) states of the theory. We make the standard assumptions that Ω is finite dimensional and compact. While Ω naturally lives inside an affine space, $\mathsf{AffSpan}[\Omega]$, for convenience, we represent it as living inside a real inner product space $(V, \langle _, _ \rangle)$ of one dimension higher, where we embed $\mathsf{AffSpan}[\Omega]$ as a hyperplane in V that does not intersect with the origin $\mathbf{0}$. This is analogous to embedding the Bloch-Ball within the real vector space of Hermitian matrices. The reason for doing so is that we can then define both the GPT states and GPT effects within the same space.

Note that we will *not* restrict attention to GPTs satisfying the no-restriction hypothesis [25], which stipulates that all the states and effects that are logically possible must also be physically possible.

A GPT also associates to every system a set of GPT *effect vectors*, $\mathcal{E}$. In the framework of GPTs, the probability of obtaining an effect $\mathbf{e} \in \mathcal{E}$ given a state $\mathbf{s} \in \Omega$ is given by the inner product:

$$\mathrm{Prob}(\mathbf{e}, \mathbf{s}) := \langle \mathbf{e}, \mathbf{s} \rangle. \tag{1}$$

We require that $\mathcal{E}$ must satisfy the following constraints. If one defines the dual of Ω, denoted Ω^*, as the set of vectors in V whose inner product with all state vectors in Ω is between 0 and 1, i.e.,

$$\Omega^* := \{\mathbf{x} \in V \mid \langle \mathbf{x}, \mathbf{s} \rangle \in [0, 1] \text{ for all } \mathbf{s} \in \Omega\}, \tag{2}$$

then $\mathcal{E}$ is a compact convex set contained in Ω^*, $\mathcal{E} \subseteq \Omega^*$, which contains the origin $\mathbf{0}$ and the "unit effect" $\mathbf{u}$, which in turn respectively satisfy $\langle \mathbf{0}, \mathbf{s} \rangle = 0$ and $\langle \mathbf{u}, \mathbf{s} \rangle = 1$ for all $\mathbf{s} \in \Omega$. Because of how we embedded $\mathsf{AffSpan}[\Omega]$ within V, $\mathbf{u}$ necessarily exists and is unique [26].

The state and effect spaces of any valid GPT must satisfy the principle of *tomography*, which states that the GPT states and GPT effects can be uniquely identified by the probabilities that they produce. Formally, for the GPT states, we have $\langle \mathbf{e}, \mathbf{s_1} \rangle = \langle \mathbf{e}, \mathbf{s_2} \rangle$ for all $\mathbf{e} \in \mathcal{E}$ if and only if $\mathbf{s_1} = \mathbf{s_2}$, and for the GPT effects, we have $\langle \mathbf{e_1}, \mathbf{s} \rangle = \langle \mathbf{e_2}, \mathbf{s} \rangle$ for all $\mathbf{s} \in \Omega$ if and only if $\mathbf{e_1} = \mathbf{e_2}$.
A GPT G, therefore, is defined as a quadruple $G := (V, \langle _, _ \rangle, \Omega, \mathcal{E})$ satisfying these constraints.

IV. THE GPT ASSOCIATED TO AN OPERATIONAL THEORY

The GPT associated to an operational theory T is the theory that one obtains when one quotients T relative to the notion of operational equivalence defined above. It is specified by a pair of quotienting maps

$$\mathbf{s}_ : \mathsf{Preps} \to \Omega \quad \text{and} \quad \mathbf{e}_ : \mathsf{Effects} \to \mathcal{E} \tag{3}$$

taking each preparation P (operational effect $[k|M]$) to a GPT state vector $\mathbf{s}_P$ (GPT effect vector $\mathbf{e}_{[k|M]}$) representing its operational equivalence class. The maps jointly satisfy the constraint that

$$\Pr([k|M], P) = \langle \mathbf{e}_{[k|M]}, \mathbf{s}_P \rangle \tag{4}$$

for all preparations and effects in the operational theory.

Note that Eq. (4) and the assumption of tomography guarantee that every operationally equivalent pair of preparations (effects) in the operational theory is mapped to the same GPT state (GPT effect) vector, and hence that each GPT vector is a representation of an operational equivalence class of operational procedures. That is,

$$P \simeq P' \quad \Longleftrightarrow \quad \mathbf{s}_P = \mathbf{s}_{P'} \tag{5}$$

$$\text{and} \quad [k|M] \simeq [k'|M'] \quad \Longleftrightarrow \quad \mathbf{e}_{[k|M]} = \mathbf{e}_{[k'|M']}. \tag{6}$$

Furthermore, they imply that nontrivial convex and coarse-graining relations holding among preparations (respectively effects) in the operational theory are encoded in the geometric relations between the GPT state vectors (respectively GPT effect vectors) in the GPT. For example, if P_1 is a convex mixture of P_2 and P_3 with weights w and $1 - w$ (or if P_1 is operationally equivalent to such a mixture), then it follows that $\mathbf{s}_{P_1} = w\mathbf{s}_{P_2} + (1 - w)\mathbf{s}_{P_3}$.

V. ONTOLOGICAL MODEL OF AN OPERATIONAL THEORY

An *ontological model* of an operational theory T is an attempt to provide a causal explanation of the operational statistics of T. For a prepare-measure experiment, it posits that the response of the measurement is determined (possibly probabilistically) by the ontic state λ of the system (a complete characterization of its physical attributes), while preparation procedures determine the distribution over the space of ontic states, Λ, from which λ is sampled.

More precisely, an ontological model associates to each preparation $P \in \mathsf{Preps}$ a normalized probability distribution over Λ, denoted μ_P, representing an agent's knowledge of the ontic state when they know that the preparation was P. Denoting the set of such distributions by $\mathcal{D}[\Lambda]$, the ontological model specifies a map

$$\mu_ : \mathsf{Preps} \to \mathcal{D}[\Lambda]. \tag{7}$$

Furthermore, an ontological model associates to each operational effect $[k|M] \in \mathsf{Effects}$ a response function on Λ,

denoted $\xi_{[k|M]}$, where $\xi_{[k|M]}(\lambda)$ represents the probability assigned to the outcome k in a measurement of M if the ontic state of the system fed into the measurement device were known to be $\lambda \in \Lambda$. Denoting the set of such response functions by $\mathcal{F}[\Lambda]$, the ontological model specifies a map

$$\xi_{_} : \mathsf{Effects} \to \mathcal{F}[\Lambda]. \tag{8}$$

These two maps must preserve the convex and coarse-graining relations between operational procedures that were discussed above. For example, if P_1 is a convex mixture of P_2 and P_3 with weights w and $1-w$, then $\mu_{P_1} = w\mu_{P_2} + (1-w)\mu_{P_3}$ [3], and similarly for operational effects. Finally, the ontological model must reproduce the probability rule of the operational theory T via

$$\Pr([k|M], P) = \sum_{\lambda\in\Lambda} \xi_{[k|M]}(\lambda)\mu_P(\lambda) \tag{9}$$

(where we have assumed Λ to be discrete for simplicity).

Generalized noncontextuality.— We are now in a position to define the notion of classicality of an operational theory T with which we are concerned in this article, namely, the existence of a generalized-noncontextual ontological model of T. An ontological model of a prepare-measure experiment satisfies generalized noncontextuality if every two procedures that are operationally equivalent have identical representations in the ontological model.

In other words, the constraint for preparations is that

$$P \simeq P' \quad \Longrightarrow \quad \mu_P = \mu_{P'}, \tag{10}$$

while the constraint for operational effects is that

$$[k|M] \simeq [k'|M'] \quad \Longrightarrow \quad \xi_{[k|M]} = \xi_{[k'|M']}. \tag{11}$$

These constraints formalize the Leibnizian principle discussed in the introduction insofar as the empirical indiscernibility of procedures (the antecedents) imply the equality of their ontological representations (the consequents).

VI. ONTOLOGICAL MODEL OF GPT

As is the case for an ontological model of an operational theory, an ontological model of a GPT is an attempt to provide a causal explanation of the operational statistics in terms of a space Λ of ontic states for the system. In this case, however, what is being modeled ontologically are not preparations and measurements, but operational equivalence classes thereof. Thus, an ontological model of a GPT associates to each GPT state vector $\mathbf{s} \in \Omega$ a normalized probability distribution over Λ, denoted $\tilde{\mu}_{\mathbf{s}} \in \mathcal{D}[\Lambda]$, and to each GPT effect vector $\mathbf{e} \in \mathcal{E}$ a response function on Λ, denoted $\tilde{\xi}_{\mathbf{e}} \in \mathcal{F}[\Lambda]$. Hence, it specifies a pair of maps

$$\tilde{\mu}_{_} : \Omega \to \mathcal{D}[\Lambda] \quad \text{and} \quad \tilde{\xi}_{_} : \mathcal{E} \to \mathcal{F}[\Lambda], \tag{12}$$

which must be linear by the assumption that they preserve the convex and coarse-graining relations defined by the geometry of the GPT state and GPT effect spaces. Finally, the ontological model must reproduce the probability rule of the GPT via

$$\langle \mathbf{e}, \mathbf{s} \rangle = \sum_{\lambda\in\Lambda} \tilde{\xi}_{\mathbf{e}}(\lambda)\tilde{\mu}_{\mathbf{s}}(\lambda). \tag{13}$$

It is now clear that a generalized-noncontextual ontological model of an operational theory T is equivalent to an ontological model of the GPT associated to T. (An explicit proof is given in Appendix A.) Hence, we have the following result.

Proposition 1. *There exists a generalized-noncontextual ontological model of an operational theory T describing prepare-measure experiments on a system if and only if there exists an ontological model of the GPT G that T defines.*

As we show in Appendix B, an ontological model of a GPT is equivalent to a positive quasiprobability representation of that GPT. Hence, Proposition 1 is the generalization of the results of Refs. [6,7] from quantum theory to an arbitrary GPT.

Note that ontological models of GPTs, unlike those of operational theories, cannot be said to be either generalized-contextual or generalized-noncontextual. Recall that contexts are defined as differences among procedures that are operationally equivalent, so there is no notion of context in a GPT, since the latter is obtained by quotienting relative to operational equivalences. To ask whether the ontological representation of a GPT state (or GPT effect) varies with context is a category mistake since there is no variability of context for GPT states (or GPT effects), just as it is a category mistake to ask whether X varies with Y when Y exhibits no variability [27].

This implies another contrast between ontological models of GPTs and those of operational theories. For a given operational theory, one can *always* construct an ontological model by allowing this model to be generalized contextual (using the analogue of the construction of Ref. [28]). But it is *not* the case that one can always construct an ontological model of a GPT because such models do not have the benefit of the representational flexibility afforded by nontrivial context dependences. Indeed, it is this lack of flexibility that implies the necessity of negativity in quasiprobability representations of some GPTs [6,7].

VII. THE GEOMETRIC CRITERION ASSOCIATED TO NONCONTEXTUALITY

We argued in the introduction that an operational theory is best viewed as classical if its operational predictions admit of an explanation in terms of a generalized-noncontextual ontological model. It is natural, therefore, to determine what this notion of classicality for an operational theory entails for the GPT that the latter defines. By Proposition 1, this is equivalent to finding a criterion for when a GPT admits of an ontological model. We now give such a condition, which we term simplex embeddability, under the assumption that the ontological model has an ontic state space Λ of finite cardinality. At the end of the article, we mention some followup works that lift this restriction.

Definition 1 (Simplex-embeddable GPTs)**.** *A GPT describing a prepare-measure experiment, $G = (V, \langle _, _ \rangle_V, \Omega, \mathcal{E})$, is simplex-embeddable if and only if there exists (i) an inner product space $(W, \langle _, _ \rangle_W)$ of some dimension d that contains a $(d-1)$-dimensional (hence d-vertex) simplex Δ_d (whose affine span does not contain the origin) and its dual hypercube Δ_d^*, and (ii) a pair of linear maps $\iota, \kappa : V \to W$ satisfying*

$$\iota(\Omega) \subseteq \Delta_d, \tag{14}$$

$$\kappa(\mathcal{E}) \subseteq \Delta_d^*, \tag{15}$$

$$\langle \boldsymbol{e}, \boldsymbol{s} \rangle_V = \langle \kappa(\boldsymbol{e}), \iota(\boldsymbol{s}) \rangle_W \quad \textit{for all } \boldsymbol{e} \in \mathcal{E} \textit{ and all } \boldsymbol{s} \in \Omega. \tag{16}$$

Note that while it is only the space of GPT *states* that embeds within a simplex, while the space of GPT effects embeds within a hypercube dual to this simplex, we nonetheless use the term "simplex-embeddable" as an umbrella term for the pair of embedding relations.

With this definition in hand, one can prove the following result (as we show in Appendix C).

Theorem 2. *A GPT describing a prepare-measure experiment admits of an ontological model over an ontic space Λ of finite cardinality if and only if it is simplex-embeddable.*

Crucially, note that the dimension of the vector space in which this embedding can be constructed may be greater than the native dimension of the GPT. We provide an explicit example of the necessity of such a "dimension gap" in Appendix D.

By combining Proposition 1 and Theorem 2, one immediately obtains our main result, Theorem 1, which gives a geometric characterization of the set of GPTs that are associated to operational theories that admit of generalized-noncontextual ontological models over ontic state spaces of finite cardinality.

VIII. DISCUSSION

As noted earlier, the prevailing view up to now has been that a GPT should be deemed classical if and only if it is simplicial, i.e., if its state space is a simplex and its effect space is the hypercube that is dual to this simplex. To understand the distinction between simpliciality and simplex embeddability, we recall a distinction between notions of nonclassicality introduced in Refs. [29,30]: an operational theory is deemed weakly nonclassical if it exhibits measurement incompatibility (sets of measurements that cannot all be simulated by processing the outcome of a single measurement) or ambiguity of mixtures (mixed states with multiple convex decompositions into pure states) or both, while it is deemed strongly nonclassical if it furthermore fails to admit of a generalized-noncontextual ontological model. This distinction can be extended from operational theories to GPTs using Proposition 1: a GPT is deemed *weakly nonclassical* if it exhibits measurement incompatibility [31] or ambiguity of mixtures or both, while it is deemed *strongly nonclassical* if it furthermore fails to admit of an ontological model. Because a GPT exhibits measurement incompatibility if and only if its effect space is not a hypercube [32] and it exhibits ambiguity of mixtures if and only if its state space is not a simplex, it follows that the notion of nonclassicality captured by nonsimpliciality of a GPT (the focus of previous work) is merely weak nonclassicality. By contrast, the notion of nonclassicality captured by the failure of *simplex embeddability* (introduced here) is exactly that of strong nonclassicality.

For both operational theories and GPTs, the strong notion of nonclassicality captures the idea that operational predictions *resist a classical explanation*. In light of Theorem 1, each of the motivations listed in the introduction for taking generalized noncontextuality as a good notion of classical explainability *for operational theories* can be reappropriated as a motivation for taking simplex embeddability as a good notion of classical explainability *for a GPT*. In the reverse direction, Theorem 1 provides a novel and independently motivated justification for generalized noncontextuality as a good notion of classical explainability of an operational theory, since at the level of the GPT, simplex embeddability is a very natural way to formalize the notion of classical explainability.

In related work undertaken simultaneously, Shahandeh [33] considered the consequences of a notion of noncontextuality, termed *broad noncontextuality*, which differs from the notion of generalized noncontextuality. As we show in the Appendix E, our results imply that the geometric condition on a GPT associated to this notion is

simplex embeddability *without a dimension gap*. Shahandeh [33] focused on GPTs that satisfy the no-restriction hypothesis [25] and demonstrated that, for such GPTs, simplex embeddability without a dimension gap coincides with simpliciality [34]. We discuss the relation between Ref. [33] and this work further in the Appendix E, and also demonstrate that it is *only* for GPTs satisfying the no-restriction hypothesis that simplex embeddability and simpliciality coincide. Barnum and Lami [35] found similar results to those presented here and in Ref. [33], while also considering generalizations to infinite dimensions.

IX. APPLICATIONS

Our result provides a novel way to test, for a given set of experimentally realized preparations and measurements that are tomographically complete, whether or not these provide evidence of strong nonclassicality. One first determines the set of GPTs that are compatible with the data obtained from prepare-measure experiments, using the techniques described in Ref. [36] (because of noise and finite precision effects, it is never a *single* GPT that is picked out by the data). One then tests these GPTs for simplex embeddability.

The previous gold standard for testing generalized noncontextuality [21,37,38] was to test a specific noise-robust noncontextuality inequality, obtained from a specific set of operational equivalences [39]. Such a test, however, requires an experimentalist to target a set of preparations and a set of measurements satisfying these specific operational equivalences [40]. The simplex-embedding test for generalized noncontextuality, by contrast, makes use of *whatever* operational equivalences happen to be satisfied by the preparations and measurements that were physically realized. It also makes use of *all* such equivalences, since these relations are encoded in the geometry of the state and effect spaces. Consequently, one need not design an experiment to target particular equivalences [41]. Rather, the technique can be applied to data obtained in *any* experiment that achieves tomographic completeness, including those not dedicated to testing noncontextuality. The technique's scope of applicability is therefore much greater and consequently promises greater applications. For instance, our method can be used to determine whether prepare-measure data obtained in some experimental architecture for quantum computation (e.g., in a noisy implementation of some simple quantum algorithm) either does or does not witness the presence of strong nonclassicality.

It is worth noting two practical issues with testing for simplex embeddability. Firstly, one needs some algorithm for testing whether a generic GPT G (described geometrically) admits an embedding of the type described in Definition 1 for a given dimension d. Secondly, even given such an algorithm, the question arises of whether there is an *upper bound* on the dimension d up to which one must apply this test. The latter question has been addressed by subsequent work, described below.

Note that to merely witness nonclassicality of a given GPT, it suffices to find inner approximations of its state and effect spaces that do not embed in a simplex and its dual. Similarly, to merely witness classicality, it suffices to find outer approximations of these that *do* embed in a simplex and its dual. The problem of finding such witnesses can be simplified, therefore, by making a propitious choice of the shape of these inner or outer approximations. This trick can also be applied to experimental data—it suffices to choose convenient inner and outer approximations to the spaces of *all* the GPTs in the set compatible with the data.

Subsequent work.— Two of the key questions raised by this work have been resolved by subsequent works. First, the results of Refs. [42] and [43] imply that the dimensional caveats in our main theorems can be dropped, removing the primary obstacle to testing for simplex embeddability in practice. In particular, Corollary 33 of Ref. [42] provides a dimension bound given by the square of the GPT's dimension. For tomographically local [44] GPTs, Ref. [43] provided a tight bound given by the GPT's dimension itself (under some natural assumptions). (It follows that every GPT of finite dimension that admits of an ontological model admits of one with a finite number of ontic states.) Second, Ref. [43] extended our result beyond the prepare-measure scenario to general compositional scenarios. (Ref. [45] then leveraged this compositional approach to prove that there is a unique noncontextual representation for every odd-dimensional stabilizer subtheory.)

ACKNOWLEDGMENTS

We thank Matt Pusey and Thomas Galley for helpful discussions. R.W.S. thanks Howard Barnum, Alex Wilce, and Jonathan Barrett for early discussions on the topic of this article. D.S. is supported by a Vanier Canada Graduate Scholarship. J.H.S. is supported in part by the Foundation for Polish Science through IRAP project cofinanced by the EU within Smart Growth Operational Programme (Contract No. 2018/MAB/5). R.K. is supported by the Chargé des recherches fellowship of the Fonds de la Recherche Scientifique - FNRS (F.R.S.-FNRS), Belgium. This research is supported by the Perimeter Institute for Theoretical Physics. Research at the Perimeter Institute is supported in part by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Colleges and Universities.

APPENDIX A: PROOF OF PROPOSITION 1

In one direction, given an ontological model of the GPT G that is associated to the operational theory T (via the quotienting maps $\mathbf{s}_$ and $\mathbf{e}_$), i.e., given the relevant maps $\tilde{\mu}_$ and $\tilde{\xi}_$, we can construct a generalized-noncontextual

ontological model for T by simply composing the quotienting map followed by the ontological map; that is, by constructing

$$\mu_{_} := \tilde{\mu}_{\mathbf{s}_{_}} \tag{A1}$$

and

$$\xi_{_} := \tilde{\xi}_{\mathbf{e}_{_}}. \tag{A2}$$

It is then easy to check that these maps define an ontological model of the operational theory T. Firstly, $\mu_{_}$ preserves the convex relations among preparations, since $\mathbf{s}_{_}$ preserves these relations and $\tilde{\mu}_{_}$ is linear. Similarly, $\xi_{_}$ preserves the convex and coarse-graining relations for operational effects. The resulting model is easily seen to be noncontextual, since

$$P_1 \simeq P_2 \quad \Longrightarrow \quad \mathbf{s}_{P_1} = \mathbf{s}_{P_2}, \tag{A3}$$

$$\Longrightarrow \quad \tilde{\mu}_{\mathbf{s}_{P_1}} = \tilde{\mu}_{\mathbf{s}_{P_2}}, \tag{A4}$$

$$\Longrightarrow \quad \mu_{P_1} = \mu_{P_2}, \tag{A5}$$

and similarly for operational effects. Finally, the two maps together reproduce the predictions of the operational theory, since

$$\begin{aligned}\sum_{\lambda\in\Lambda} \xi_{[k|M]}(\lambda)\mu_P(\lambda) &= \sum_{\lambda\in\Lambda} \tilde{\xi}_{\mathbf{e}_{[k|M]}}(\lambda)\tilde{\mu}_{\mathbf{s}_P}(\lambda)\\ &\overset{(13)}{=} \langle \mathbf{e}_{[k|M]}, \mathbf{s}_P\rangle\\ &\overset{(4)}{=} \Pr([k|M], P).\end{aligned} \tag{A6}$$

Conversely, given a generalized-noncontextual ontological model of an operational theory T (i.e., given the relevant maps $\mu_{_}$ and $\xi_{_}$), we can construct an ontological model of the GPT G associated to it (via the relevant maps $\mathbf{s}_{_}$ and $\mathbf{e}_{_}$) by defining the maps $\tilde{\mu}_{_}$ and $\tilde{\xi}_{_}$ as the unique linear maps satisfying

$$\tilde{\mu}_{\mathbf{s}} := \mu_P \tag{A7}$$

for all preparations P such that $\mathbf{s}_P = \mathbf{s}$, and

$$\tilde{\xi}_{\mathbf{e}} := \xi_{[k|M]} \tag{A8}$$

for all operational effects $[k|M]$ such that $\mathbf{e}_{[k|M]} = \mathbf{e}$. It is then easy to check that $\tilde{\mu}_{_}$ preserves the convex relations between state vectors. For example, suppose that P_1 is a convex mixture of P_2 and P_3 with weights w and $1-w$, so that $\mu_{P_1} = w\mu_{P_2} + (1-w)\mu_{P_3}$ (since $\mu_{_}$ preserves the convex relations among preparation procedures). We therefore find that $\tilde{\mu}_{\mathbf{s}_{P_1}} = \mu_{P_1} = w\mu_{P_2} + (1-w)\mu_{P_3} = w\tilde{\mu}_{\mathbf{s}_{P_2}} + (1-w)\tilde{\mu}_{\mathbf{s}_{P_3}}$. Thus, in the example under consideration, $\mathbf{s}_{P_1} = w\mathbf{s}_{P_2} + (1-w)\mathbf{s}_{P_3}$ (given the operational equivalence of P_1 with the mixture of P_2 and P_3) and $\tilde{\mu}_{\mathbf{s}_{P_1}} = w\tilde{\mu}_{\mathbf{s}_{P_2}} + (1-w)\tilde{\mu}_{\mathbf{s}_{P_3}}$, and so $\tilde{\mu}_{_}$ does indeed preserve the convex relations among GPT states. The proof that $\tilde{\xi}_{_}$ preserves the convex and coarse-graining relations among GPT effects is analogous.

Finally, the two maps together reproduce the predictions of the GPT, since

$$\begin{aligned}\sum_{\lambda\in\Lambda} \tilde{\xi}_{\mathbf{e}}(\lambda)\tilde{\mu}_{\mathbf{s}}(\lambda) &= \sum_{\lambda\in\Lambda} \xi_{[k|M]}(\lambda)\mu_P(\lambda)\\ &\overset{(9)}{=} \Pr(P, [k|M])\\ &\overset{(4)}{=} \langle \mathbf{e}_{[k|M]}, \mathbf{s}_P\rangle\\ &= \langle \mathbf{e}, \mathbf{s}\rangle.\end{aligned} \tag{A9}$$

This completes the proof.

APPENDIX B: QUASIPROBABILISTIC REPRESENTATIONS OF GPTs

We have argued that a GPT should be deemed classical if and only if an ontological representation of it exists. We now prove the following proposition, which implies that this notion of classicality is equivalent to another notion of classicality, namely, the existence of a positive quasiprobability representation.

Proposition 2. *An ontological model of a GPT G describing prepare-measure experiments on a system is equivalent to a positive quasiprobability representation thereof.*

A quasiprobability representation of a GPT associates to each system a set Λ. It associates to each GPT state on this system, $\mathbf{s} \in \Omega$, a real-valued function over Λ, denoted $\hat{\mu}_{\mathbf{s}}$ and satisfying $\sum_{\lambda\in\Lambda} \hat{\mu}_{\mathbf{s}}(\lambda) = 1$. This is termed a *quasiprobability distribution* over Λ because if the function were valued in $[0, 1]$ rather than the reals, it could be interpreted as a probability distribution over Λ. Formally, it specifies a linear map from state vectors to the real vector space of functions from Λ to $\mathbb{R}$, denoted $\mathbb{R}^\Lambda$. That is,

$$\hat{\mu}_{_} : \Omega \to \mathbb{R}^\Lambda. \tag{B1}$$

It also associates to each GPT effect $\mathbf{e}$ a real-valued function over Λ, denoted $\hat{\xi}_{\mathbf{e}}$ and satisfying $\sum_{\mathbf{e}\in\chi} \hat{\xi}_{\mathbf{e}}(\lambda) = 1$ for all $\lambda \in \Lambda$, where χ is any set of GPT effects corresponding to the set of outcomes of a possible measurement. Note that such a set must satisfy $\sum_{\mathbf{e}\in\chi} \mathbf{e} = \mathbf{u}$. The function $\hat{\xi}_{\mathbf{e}}$ is termed a *quasiresponse function* because if the function were valued in $[0, 1]$ rather than the reals, it could be interpreted as a response function. Formally, it specifies a linear map

$$\hat{\xi}_{_} : \mathcal{E} \to \mathbb{R}^\Lambda. \tag{B2}$$

Now consider the probability of obtaining GPT effect $\mathbf{e}$ given a preparation associated to state $\mathbf{s}$, which is given by $\langle \mathbf{e}, \mathbf{s}\rangle_V$ in the GPT. In a quasiprobability representation of this GPT, one computes this probability using the same formula that would be appropriate if the quasiprobabilities are true probabilities, namely, $\sum_{\lambda\in\Lambda} \hat{\xi}_{\mathbf{e}}(\lambda)\hat{\mu}_{\mathbf{s}}(\lambda)$. Thus, for all $\mathbf{e} \in \mathcal{E}$ and all $\mathbf{s} \in \Omega$,

$$\langle \mathbf{e}, \mathbf{s}\rangle_V = \hat{\xi}_{\mathbf{e}} \cdot \hat{\mu}_{\mathbf{s}} = \sum_{\lambda\in\Lambda} \hat{\xi}_{\mathbf{e}}(\lambda)\hat{\mu}_{\mathbf{s}}(\lambda). \tag{B3}$$

Finally, we say that a quasiprobability representation of a GPT is *positive* if, for all $\mathbf{s} \in \Omega$ and all $\lambda \in \Lambda$, $0 \le \hat{\mu}_{\mathbf{s}}(\lambda) \le 1$, and for all $\mathbf{e} \in \mathcal{E}$ and all $\lambda \in \Lambda$, $0 \le \hat{\xi}_{\mathbf{e}}(\lambda) \le 1$.

Proposition 2 follows immediately by noting that Eqs. (B1) and (B2) together with the positivity constraints are equivalent to Eq. (12), and that Eq. (B3) is equivalent to Eq. (13).

APPENDIX C: PROOF OF THEOREM 2

It is convenient for the proof to formally define the notion of a *simplicial GPT* that was discussed in the main text. A simplicial GPT of given dimension d is a tuple $(W, \langle _,_\rangle_W, \Delta_d, \Delta_d^*)$ defined by a vector space W with inner product $\langle _,_\rangle_W$, a simplicial state space $\Omega = \Delta_d$ (of intrinsic dimension $d-1$), and the dual hypercube $\mathcal{E} = \Delta_d^*$ (of intrinsic dimension d) as its effect space. Clearly, the inner product space, simplex, and the dual appearing in the definition of simplex embeddability (Definition 1) define such a simplicial GPT [46].

Next, it is useful to characterize the inherent degeneracies in the definition of a GPT, that is, to characterize under what conditions two GPTs make the same operational predictions, and hence should be deemed equivalent.

Definition 2 (Equivalent GPTs)**.** *Two GPTs, $G = (V, \langle _,_\rangle_V, \Omega, \mathcal{E})$ and $G' = (V', \langle _,_\rangle_{V'}, \Omega', \mathcal{E}')$, are said to be equivalent if and only if there exist linear isomorphisms $\omega, \epsilon : V \to V'$ such that*

$$\omega(\Omega) = \Omega', \tag{C1}$$

$$\epsilon(\mathcal{E}) = \mathcal{E}', \tag{C2}$$

$$\langle \boldsymbol{e}, \boldsymbol{s}\rangle_V = \langle \epsilon_{\boldsymbol{e}}, \omega_{\boldsymbol{s}}\rangle_{V'} \quad \textit{for all } \boldsymbol{e} \in \mathcal{E}, \boldsymbol{s} \in \Omega. \tag{C3}$$

As a simple example, note that the GPT $(V, \langle _,_\rangle_V, \Omega, T(\mathcal{E}))$, where T is a reversible map, is equivalent to the GPT $(V, \langle _,_\rangle_V, T^\dagger(\Omega), \mathcal{E})$ where "$\dagger$" denotes the adjoint operation relative to the inner product on the vector space [22].

Now, note that any simplicial GPT of a given dimension d is equivalent (in the above sense) to one in a particular canonical form, defined as follows. The vector space is taken to be $\mathbb{R}^d$ with some chosen orthonormal basis $\{\mathbf{b}_i\}_{i\in B}$ and the inner product is taken to be the dot product with respect to this basis. The state space is the unit simplex (with d vertices and intrinsic dimension $d-1$), defined as

$$\Delta_d := \mathsf{ConvexHull}[\{\mathbf{b}_i\}], \tag{C4}$$

and the effect space is the dual of it, namely the unit hypercube (with 2^{d-1} vertices and intrinsic dimension d)

$$\Delta_d^* = \mathsf{ConvexHull}\left[\left\{\sum_{i\in\chi} \mathbf{b}_i \middle| \chi \subseteq B\right\}\right]. \tag{C5}$$

Finally, the unit effect is $\mathbf{u} = \sum_{i\in B} \mathbf{b}_i$.

At this point it is straightforward to show that the mathematical structures defining a simplicial GPT are in one-to-one correspondence with the structures defining an ontological theory.

First, one identifies one ontic state $\lambda \in \Lambda$ with each vertex $i \in \mathsf{Vert}[\Delta_d]$; that is, one makes a one-to-one correspondence $\lambda \in \Lambda \leftrightarrow i \in \mathsf{Vert}[\Delta_d]$.

Next, one can identify the GPT vector space V with the vector space $\mathbb{R}^\Lambda$ of functions from Λ to $\mathbb{R}$, where the function associated with each GPT vector is simply its coordinates along each element of the chosen basis; that is, there is a one-to-one correspondence $\mathbb{R}^\Lambda \leftrightarrow V$. Explicitly, any vector $\mathbf{w} \in V$ is uniquely identified with a function $f_{\mathbf{w}} \in \mathbb{R}^{\mathsf{Vert}[\Delta_d]}$ by $f_{\mathbf{w}}(i) := \langle \mathbf{w}, \mathbf{b}_i\rangle_V$, where $\mathbf{b}_i$ is the orthonormal basis used in the definition of the canonical form (it is the set of vectors that point to the vertices of Δ_d). Clearly, then, the vertices of Δ_d will be represented by point distributions in $\mathbb{R}^{\mathsf{Vert}[\Delta_d]}$ as $f_{\mathbf{b}_j}(i) = \langle \mathbf{b}_j, \mathbf{b}_i\rangle_V = \delta_{ij}$. By linearity, the other vectors in Δ_d will correspond to mixtures of these, and hence to probability distributions over $\mathsf{Vert}[\Delta_d]$. As such, there is a one-to-one correspondence $\mathcal{D}[\Lambda] \leftrightarrow \Delta_d$.

Similarly, the vectors in Δ_d^* can be identified with the response functions. The extremal points of Δ_d^*, namely, $\sum_{i\in\chi} \mathbf{b}_i$, are represented by the functions $f_{\sum_{j\in\chi} \mathbf{b}_j}(i) = \sum_{j\in\chi} \langle \mathbf{b}_j, \mathbf{b}_i\rangle_V = \sum_{j\in\chi} \delta_{ij}$, and these correspond to the extremal response functions in $\mathcal{F}[\Lambda]$. Again by linearity, we see that every vector in Δ_d^* is identified with a valid response function. Additionally, the vector $\mathbf{u}$ corresponds to a vector $\mathbf{1}$ whose components are all equal to 1. Hence, we have the one-to-one correspondences $F[\Lambda] \leftrightarrow \Delta_d^*$ and $\mathbf{1} \leftrightarrow \mathbf{u}$.

Finally, the dot product can now be written out in terms of these functions as $\langle \mathbf{w}, \mathbf{w}'\rangle_V = \sum_{i\in\mathsf{Vert}[\Delta_d]} f_{\mathbf{w}}(i) f_{\mathbf{w}'}(i)$, where the right-hand side defines the dot product in the ontological space. We denote this correspondence as $\sum_{\lambda\in\Lambda} f_{_}(\lambda) f_{_}(\lambda) \leftrightarrow \langle _,_\rangle_V$. We summarize the above in Table I.

Hence, any simplex-embeddable GPT G can be embedded into a simplicial GPT, whose canonical form together

TABLE I. A summary of the correspondences between ontological and geometric concepts.

Ontological	Geometric
$\lambda \in \Lambda$	$i \in \mathsf{Vert}[\Delta_d]$
$\mathbb{R}^\Lambda$	V
$\mathcal{D}[\Lambda]$	Δ_d
$\mathcal{F}[\Lambda]$	Δ_d^*
$\mathbf{1}$	$\mathbf{u}$
$\sum_{\lambda\in\Lambda} f_{_}(\lambda) f_{_}(\lambda)$	$\langle _, _ \rangle_V$

with the identification in the table can be used to explicitly construct an ontological model of G.

Conversely, any GPT that admits of an ontological model can be embedded into a simplex and its dual, which can be explicitly constructed using the identifications provided in Table I.

This completes the proof of Theorem 2.

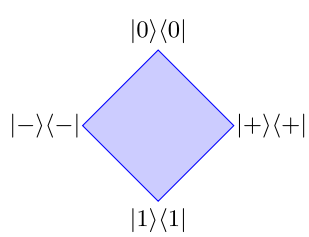
APPENDIX D: EXAMPLE OF A DIMENSION MISMATCH: STABILIZER REBIT

We now give an example of a simplex-embeddable GPT whose state and effect spaces cannot be embedded into a simplex and its dual (respectively) *in a vector space of the same dimension as the GPT itself*. The example is the GPT associated with the well-known subtheory of the stabilizer qubit theory containing only the real-amplitude states and effects, termed the *stabilizer rebit* quantum subtheory.

For the sake of obtaining a GPT representation of this theory, we conceptualize the projectors as vectors in the real vector space of Hermitian operators (as opposed to the complex vector space of arbitrary operators). The GPT state space of the stabilizer rebit is then the convex subset

$$\mathsf{ConvexHull}\,[\{|0\rangle\langle 0|, |1\rangle\langle 1|, |+\rangle\langle +|, |-\rangle\langle -|\}] \tag{D1}$$

of the full qubit state space, i.e.,



and the effect space is the convex subset

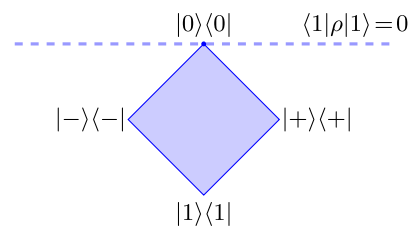
$$\mathsf{ConvexHull}[\{\mathbf{0}, |0\rangle\langle 0|, |1\rangle\langle 1|, |+\rangle\langle +|, |-\rangle\langle -|, \mathbb{1}\}] \tag{D2}$$

of the full qubit effect space.

This stabilizer rebit GPT admits of an ontological model; one such model is the toy model of Ref. [47]. (We present this model explicitly at the end of this appendix.) As demanded by Theorem 2, this implies that the stabilizer rebit state space can be embedded in a simplex whose dual contains the stabilizer rebit effect space. Indeed, the space of probability distributions over ontic states of the toy model of Ref. [47] defines a tetrahedron that contains the state space, and whose dual contains the effect space. Note that the stabilizer rebit state space is two dimensional, while the simplex (the tetrahedron) in which it is embedded is intrinsically three dimensional. As we now prove, this dimension mismatch is unavoidable: it is not possible to find an embedding into any intrinsically two dimensional simplex such that the effect space embeds in the dual.

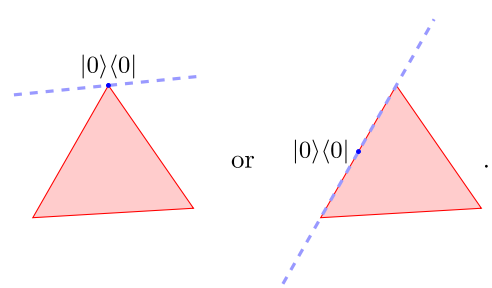
Proof. We assume that there exists a triangle that embeds the stabilizer rebit state space, whose dual embeds the stabilizer rebit effect space, and we then prove a contradiction.

Note that, for any of the extremal states, there exists an effect that evaluates to zero uniquely on that state, for example, given the state $|0\rangle\langle 0|$ then we have the effect $|1\rangle\langle 1|$ that gives probability zero only for that state. The set of states for which this effect evaluates to zero, i.e., those satisfying $\mathrm{tr}[|1\rangle\langle 1|\rho] = \langle 1|\,\rho\,|1\rangle = 0$, can be geometrically represented by a hyperplane in the state space as follows:

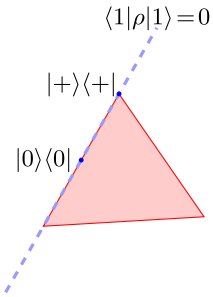


Let us consider the embedding of the state $|0\rangle\langle 0|$ into the triangle. *A priori* this could go to any point in the triangle. However, we know that there should be some effect in the dual that evaluates to zero on this state, and not on any of the other states we will embed. This means that $|0\rangle\langle 0|$ must be mapped to a point on the boundary of the triangle. (The only effect that evaluates to zero on an interior point of the triangle is the zero effect that would also evaluate to zero on all of the other states.)

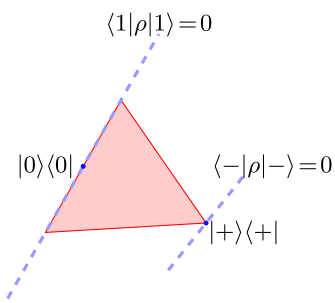
This means that $|0\rangle\langle 0|$ must lie on a proper face of the triangle, and the associated effect defines the hyperplane that picks out that face. There are two possibilities here, $|0\rangle\langle 0|$ is mapped to a vertex or an edge as follows:



Now consider a second state, e.g., $|+\rangle\langle+|$. The same argument applies again, so it must be mapped to some proper face of the triangle. Moreover, this must be disjoint from the face into which we embedded $|0\rangle\langle 0|$, since otherwise the effect associated to $|0\rangle\langle 0|$ would also evaluate to zero on $|+\rangle\langle+|$, as in the following example:



That is, it would imply that $\mathrm{tr}[|1\rangle\langle 1|\,|+\rangle\langle+|] = 0$, and so the ontological model would give predictions different from those of the GPT. Hence, $|0\rangle\langle 0|$ and $|+\rangle\langle+|$ must be mapped to disjoint faces, as in the following example:



We can apply this same reasoning to every pair of the four extremal states of the stabilizer rebit to conclude that these states must all be mapped to disjoint faces of the triangle.

But there are at most three disjoint faces of a triangle, so we reach a contradiction. ■

1. Explicit ontological model for the stabilizer rebit

This ontological model is obtained directly by taking the appropriate subsets of the state and effect spaces from the toy theory of Ref. [47]. The ontic state space is defined as

$$\Lambda \equiv \{\lambda_{0+}, \lambda_{0-}, \lambda_{1+}, \lambda_{1-}\}. \tag{D3}$$

The epistemic states are given by

$$\begin{aligned}
\tilde{\mu}_{|0\rangle\langle 0|} &= \left(\tfrac{1}{2}, \tfrac{1}{2}, 0, 0\right),\\
\tilde{\mu}_{|1\rangle\langle 1|} &= \left(0, 0, \tfrac{1}{2}, \tfrac{1}{2}\right),\\
\tilde{\mu}_{|+\rangle\langle+|} &= \left(\tfrac{1}{2}, 0, \tfrac{1}{2}, 0\right),\\
\tilde{\mu}_{|-\rangle\langle-|} &= \left(0, \tfrac{1}{2}, 0, \tfrac{1}{2}\right).
\end{aligned}$$

The response functions are given by

$$\begin{aligned}
\tilde{\xi}_{|0\rangle\langle 0|} &= (1, 1, 0, 0),\\
\tilde{\xi}_{|1\rangle\langle 1|} &= (0, 0, 1, 1),\\
\tilde{\xi}_{|+\rangle\langle+|} &= (1, 0, 1, 0),\\
\tilde{\xi}_{|-\rangle\langle-|} &= (0, 1, 0, 1).
\end{aligned}$$

It is straightforward to verify that this model reproduces the operational predictions.

APPENDIX E: COMPARING SIMPLEX EMBEDDABILITY AND SIMPLICIALITY

It is interesting to consider under which conditions the notions of simplex embeddability and simpliciality coincide—that is, when the notions of weak and strong nonclassicality coincide. Intuitively, this is connected to the no-restriction hypothesis, which states that every logically possible effect is a valid physical effect and similarly that every logically possible state is a valid physical state. This is because if a GPT is simplex-embeddable but not itself simplicial, then the extra vectors in the simplex (or its dual) constitute logically but not physically possible states (effects), and so the GPT cannot satisfy the no-restriction hypothesis.

If one assumes no dimension gap (i.e., the simplex embedding is into an inner product space of the same dimension as the GPT vector space), then this intuition about the connection between the no-restriction hypothesis and weak and strong notions of classicality can be proved in a particularly simple manner.

Theorem 3. *If a GPT $G' = (V, \langle_,_\rangle_V, \Omega', \mathcal{E}')$ satisfies the no-restriction hypothesis (that is, $\mathcal{E}' = \Omega'^*$) and is moreover simplex-embeddable using an inner product space $(W, \langle_,_\rangle_W)$ of the same dimension as V, then G' is simplicial.*

Proof. Consider the simplex embedding of a GPT $G' = (V, \langle_,_\rangle_V, \Omega', \mathcal{E}')$ using an inner product space $(W, \langle_,_\rangle_W)$ via linear maps ι and κ, such that $\iota(\Omega') \subseteq \Delta$ and $\kappa(\mathcal{E}') \subseteq \Delta^*$. As we are assuming that the dimension of W is the same as the dimension of V, it is clear that the GPT $G := (W, \langle_,_\rangle_W, \iota(\Omega'), \kappa(\mathcal{E}'))$ is equivalent to G' in the sense of Definition 2. For the remainder of this proof, therefore, we work with this equivalent GPT, G, and notate $\Omega := \iota(\Omega')$ and $\mathcal{E} := \kappa(\mathcal{E}')$. Since G' satisfies the no-restriction hypothesis (by assumption), so does G.

By the assumption of simplex embeddability, the set Ω of all states of G is contained within a simplex, $\Omega \subseteq \Delta$, and consequently every effect in the hypercube that is dual to this simplex, Δ^*, gives valid probabilities on all states in Ω. It follows that every effect in Δ^* is a logically possible effect. Under the no-restriction hypothesis, therefore, it

follows that every such effect must be physically possible, that is, $\Delta^* \subseteq \mathcal{E}$. But the assumption that G is simplex-embeddable also implies that every physically realizable effect is contained in the hypercube Δ^*, so that $\mathcal{E} \subseteq \Delta^*$. The conjunction of these gives $\mathcal{E} = \Delta^*$. By a final appeal to the no-restriction hypothesis, we find that the state space is the dual of the effect space, $\Omega = \mathcal{E}^*$, and consequently $\Omega = \Delta$. ■

Unlike the stronger result proven below, this theorem holds only for embeddings into simplicial GPTs of the same dimension as the given GPT.

This lack of a dimension gap is what was assumed in related work by Shahandeh [33], who considered the consequences of a notion termed *broad noncontextuality*. The latter differs from the notion of generalized noncontextuality insofar as the set of response functions associated to physically realizable GPT effects are assumed to be tomographically complete not just for distributions associated to physically realizable GPT states over (which is all that generalized noncontextuality demands) but for all distributions over ontic states. This additional assumption implies that any simplex embedding will be without a dimension gap.

Secondly, in Ref. [43] it was demonstrated that the lack of a dimension gap can be *derived* under certain assumptions (in particular, the assumption that the GPT is tomographically local). However, the above theorem can be strengthened to remove the assumption that there is no dimension gap. The following is an adaptation of a theorem first proved by Shahandeh in [33].

Theorem 4 (Classicality of GPTs satisfying the no-restriction hypothesis)**.** *For any GPT that satisfies the no-restriction hypothesis, that is, where $\mathcal{E} = \Omega^*$, simplex embeddability and simpliciality are equivalent.*

Proof. Using Theorem 2, we know that simplex embeddability is equivalent to the existence of an ontological model of the GPT. This ontological model is specified in terms of some linear representation maps $\tilde{\mu}_$ and $\tilde{\xi}_$. The fact that $\tilde{\mu}_$ is a linear map on Ω implies that, for each λ, $\tilde{\mu}_(\lambda)$ is a linear functional on Ω, and the fact that $\tilde{\xi}_$ is a linear map on $\mathcal{E}$ implies that, for each λ, $\tilde{\xi}_(\lambda)$ is a linear functional on $\mathcal{E}$. Riesz's representation theorem further implies that, for each λ, there must exist vectors $\mathbf{e}_\lambda$ and $\mathbf{s}_\lambda$ in V such that

$$\tilde{\mu}_(\lambda) = \langle \mathbf{e}_\lambda, _\rangle, \tag{E1}$$

$$\tilde{\xi}_(\lambda) = \langle _, \mathbf{s}_\lambda\rangle. \tag{E2}$$

The no-restriction hypothesis then implies that $\mathbf{e}_\lambda \in \mathcal{E}$ and $\mathbf{s}_\lambda \in \Omega$ for all $\lambda \in \Lambda$.

Next, note that, for any fixed $\mathbf{s} \in \Omega$, it holds that, for all $\mathbf{e} \in \mathcal{E}$,

$$\begin{aligned}\langle \mathbf{e}, \mathbf{s}\rangle &= \sum_\lambda \tilde{\mu}_\mathbf{s}(\lambda)\tilde{\xi}_\mathbf{e}(\lambda)\\ &= \sum_\lambda \tilde{\mu}_\mathbf{s}(\lambda)\langle \mathbf{e}, \mathbf{s}_\lambda\rangle\\ &= \left\langle \mathbf{e}, \sum_\lambda \tilde{\mu}_\mathbf{s}(\lambda)\mathbf{s}_\lambda\right\rangle,\end{aligned} \tag{E3}$$

and hence tomography of the GPT implies that

$$\mathbf{s} = \sum_\lambda \tilde{\mu}_\mathbf{s}(\lambda)\mathbf{s}_\lambda \tag{E4}$$

for every $\mathbf{s} \in \Omega$. Because $\tilde{\mu}_\mathbf{s}(\lambda)$ is a probability distribution, this implies that every GPT state vector $\mathbf{s}$ can be written as a convex combination of the states $\{\mathbf{s}_\lambda\}$.

Now consider some vertex $\mathbf{v} \in \mathsf{Vert}[\Omega]$, and decompose it as

$$\mathbf{v} = \sum_\lambda \tilde{\mu}_\mathbf{v}(\lambda)\mathbf{s}_\lambda. \tag{E5}$$

As $\mathbf{v}$ is a vertex (i.e., it is convexly extremal), it must be the case that, for any λ such that $\tilde{\mu}_\mathbf{v}(\lambda) \neq 0$, we have $\mathbf{s}_\lambda = \mathbf{v}$. This immediately implies that distinct vertices have ontic representations with disjoint support on the ontic state space. The proof is by contradiction. Consider distinct vertices $\mathbf{v}, \mathbf{w} \in \mathsf{Vert}[\Omega]$ and assume that their representations have overlapping support, that is, assume that there exists $\lambda^* \in \Lambda$ such that $\tilde{\mu}_\mathbf{v}(\lambda^*) \neq 0$ and $\tilde{\mu}_\mathbf{w}(\lambda^*) \neq 0$. In this case, we could infer that $\mathbf{v} = \mathbf{s}_{\lambda^*}$ and $\mathbf{w} = \mathbf{s}_{\lambda^*}$, contradicting the hypothesis that $\mathbf{v}$ and $\mathbf{w}$ are distinct.

Now let us consider two decompositions of some state $\mathbf{s}^* \in \Omega$ into vertices:

$$\sum_{\mathbf{v}\in\mathsf{Vert}[\Omega]} \alpha_\mathbf{v}\mathbf{v} = \mathbf{s}^* = \sum_{\mathbf{w}\in\mathsf{Vert}[\Omega]} \beta_\mathbf{w}\mathbf{w}. \tag{E6}$$

If the state space is not a simplex then there *must* exist nonunique decompositions. However, we now show that these two decompositions are necessarily the same, and hence that Ω is a simplex.

Applying $\tilde{\mu}$ to these two decompositions, linearity of $\tilde{\mu}$ implies that

$$\sum_{\mathbf{v}\in\mathsf{Vert}[\Omega]} \alpha_\mathbf{v}\tilde{\mu}_\mathbf{v} = \sum_{\mathbf{w}\in\mathsf{Vert}[\Omega]} \beta_\mathbf{w}\tilde{\mu}_\mathbf{w}. \tag{E7}$$

Since the $\tilde{\mu}_\mathbf{v}$ all have disjoint support, it must be that $\alpha_\mathbf{v} = \beta_\mathbf{v}$ for all $\mathbf{v} \in \mathsf{Vert}[\Omega]$, and hence the two decompositions are in fact identical. So there are no nonunique decompositions of states, and hence Ω must be a simplex.

By the assumed no-restriction hypothesis, we also infer that $\mathcal{E} = \Omega^*$, and hence the GPT is simplicial. ■

What this means is that any GPT that is simplex embeddable but not itself simplicial asserts the existence of some candidate states or some candidate effects (or some of both) that are logically possible (in the sense that they would yield valid probabilities) but stipulated to be not physically realizable. That is, such GPTs must necessarily violate the no-restriction hypothesis. The divergence of weak and strong notions of nonclassicality is therefore only manifest when we go beyond the special class of GPTs that satisfy the no-restriction hypothesis.

To show that the distinction between weak and strong nonclassicality disappears for all *and only* GPTs satisfying the no-restriction hypothesis, it remains only to establish the "only" half of the implication. This is the easy half. The only GPTs that satisfy *both* the condition of being simplicial and the condition of being simplex-embeddable are those that are simplicial, and these clearly satisfy the no-restriction hypothesis.

[1] J. S. Bell, On the Einstein Podolsky Rosen paradox, Physics **1**, 195 (1964).

[2] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Bell nonlocality, Rev. Mod. Phys. **86**, 419 (2014).

[3] R. W. Spekkens, Contextuality for preparations, transformations, and unsharp measurements, Phys. Rev. A **71**, 052108 (2005).

[4] Refs. [48–53] provided alternative proposals for how to operationalize the Kochen-Specker notion of noncontextuality [54]. See Refs. [21,55,56] for a critique of these.

[5] R. W. Spekkens, Noncontextuality: how we should define it, why it is natural, and what to do about its failure, PIRSA:17070035 (2017).

[6] R. W. Spekkens, Negativity and Contextuality are Equivalent Notions of Nonclassicality, Phys. Rev. Lett. **101**, 020401 (2008).

[7] C. Ferrie and J. Emerson, Frame representations of quantum mechanics and the necessity of negativity in quasi-probability representations, J. Phys. A **41**, 352001 (2008).

[8] M. F. Pusey, Anomalous Weak Values Are Proofs of Contextuality, Phys. Rev. Lett. **113**, 200401 (2014).

[9] R. Kunjwal, M. Lostaglio, and M. F. Pusey, Anomalous weak values and contextuality: Robustness, tightness, and imaginary parts, Phys. Rev. A **100**, 042116 (2019).

[10] R. W. Spekkens, D. H. Buzacott, A. J. Keehn, B. Toner, and G. J. Pryde, Preparation Contextuality Powers Parity-Oblivious Multiplexing, Phys. Rev. Lett. **102**, 010401 (2009).

[11] A. Chailloux, I. Kerenidis, S. Kundu, and J. Sikora, Optimal bounds for parity-oblivious random access codes, New J. Phys. **18**, 045003 (2016).

[12] A. Ambainis, M. Banik, A. Chaturvedi, D. Kravchenko, and A. Rai, Parity Oblivious *d*-Level Random Access Codes and Class of Noncontextuality Inequalities, arXiv:1607.05490 (2016).

[13] D. Saha, P. Horodecki, and M. Pawłowski, State independent contextuality advances one-way communication, New J. Phys. **21**, 093057 (2019).

[14] D. Schmid and R. W. Spekkens, Contextual Advantage for State Discrimination, Phys. Rev. X **8**, 011015 (2018).

[15] M. Lostaglio and G. Senno, Contextual advantage for state-dependent cloning, arXiv:1905.08291 (2019).

[16] M. Howard, J. Wallman, V. Veitch, and J. Emerson, Contextuality supplies the 'magic' for quantum computation, Nature **510**, 351 (2014).

[17] R. Raussendorf, Contextuality in measurement-based quantum computation, Phys. Rev. A **88**, 022322 (2013).

[18] M. J. Hoban, E. T. Campbell, K. Loukopoulos, and D. E. Browne, Non-adaptive measurement-based quantum computation and multi-party Bell inequalities, New J. Phys. **13**, 023014 (2011).

[19] S. A. Yadavalli and R. Kunjwal, Contextuality in entanglement-assisted one-shot classical communication, arXiv:2006.00469 (2020).

[20] R. W. Spekkens, The ontological identity of empirical indiscernibles: Leibniz's methodological principle and its significance in the work of Einstein, arXiv:1909.04628 (2019).

[21] M. D. Mazurek, M. F. Pusey, R. Kunjwal, K. J. Resch, and R. W. Spekkens, An experimental test of noncontextuality without unphysical idealizations, Nat. Comm. **7**, 1 (2016).

[22] L. Hardy, Quantum Theory From Five Reasonable Axioms, quant-ph/0101012 (2001).

[23] J. Barrett, Information processing in generalized probabilistic theories, Phys. Rev. A **75**, 032304 (2007).

[24] In previous work (and here), these relations are *only* stipulated in the "ordinary language" description of the procedures. In forthcoming work [57], we show how they can be incorporated into the formal structure of an operational theory.

[25] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Probabilistic theories with purification, Phys. Rev. A **81**, 062348 (2010).

[26] This uniqueness is referred to as the "causality axiom" in Ref. [25].

[27] O. Magidor, in *The Stanford Encyclopedia of Philosophy* (Metaphysics Research Lab, Stanford University, 2019).

[28] E. G. Beltrametti and S. Bugajski, A classical extension of quantum mechanics, J. Phys. A **28**, 3329 (1995).

[29] R. W. Spekkens, in *Quantum Theory: Informational Foundations and Foils*, edited by G. Chiribella and R. W. Spekkens (Springer Netherlands, Dordrecht, 2016), p. 83.

[30] Note, however, that there it was the nonclassicality of operational phenomena rather than of operational theories that was at issue.

[31] In our formalization of GPTs, a measurement is any collection of effects that sum to the unit effect.

[32] M. Plávala, All measurements in a probabilistic theory are compatible if and only if the state space is a simplex, Phys. Rev. A **94**, 042108 (2016).

[33] F. Shahandeh, Contextuality of general probabilistic theories, arXiv:1911.11059 (2020).

[34] For many years, the no-restriction hypothesis was taken to be a fundamental part of the definition of a GPT (indeed, it was only in Ref. [25] in 2009 that the term "no-restriction hypothesis" was coined), but it was later acknowledged

to be a substantial restriction on the sorts of GPTs that could be considered. Examples of GPTs that violate the no-restriction hypothesis include: the GPT associated to Spekkens' toy model [47], the stabilizer subtheories of quantum theory as well as the Gaussian and quadrature subtheories of quantum mechanics [29], and the GPTs considered in, for example, Refs. [36,58–62].

[35] H. Barnum and L. Lami, Classical embeddability of general probabilistic theories, Forthcoming.

[36] M. D. Mazurek, M. F. Pusey, K. J. Resch, and R. W. Spekkens, Experimentally bounding deviations from quantum theory in the landscape of generalized probabilistic theories, arXiv:1710.05948 (2017).

[37] M. F. Pusey, L. del Rio, and B. Meyer, Contextuality without access to a tomographically complete set, arXiv:1904.08699 (2019).

[38] D. Schmid, R. W. Spekkens, W. Robert, 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).

[39] This set is usually inferred from a preexisting no-go theorem for generalized noncontextuality in quantum theory, for instance, those inspired by Kochen-Specker contradictions [55,63–65].

[40] In fact, because the procedures that are experimentally realized always deviate from the targets, to satisfy the operational equivalences, one must shift attention to "secondary" procedures that lie within the convex hull of the realized ones, a step that compromises some purity [66].

[41] Nor postprocess the data to satisfy them.

[42] V. Gitton and M. P. Woods, Solvable Criterion for the Contextuality of any Prepare-and-Measure Scenario, arXiv:2003.06426 (2020).

[43] D. Schmid, J. H. Selby, M. F. Pusey, and R. W. Spekkens, A structure theorem for generalized-noncontextual ontological models, arXiv:2005.07161 (2020).

[44] L. Hardy, Reformulating and reconstructing quantum theory, arXiv:1104.2066 (2011).

[45] D. Schmid, H. Du, J. H. Selby, and M. F. Pusey, The only noncontextual model of the stabilizer subtheory is Gross's, arXiv:2101.06263 (2021).

[46] For a GPT G that is simplex-embeddable but not itself simplicial, the embedding relation cannot be interpreted as merely due to technological limitations. To posit that a world is governed by the GPT G is to posit that the states and effects that are physically realizable are all and only the states and effects included in G. Consequently, any states and effects that are included in the simplicial GPT into which G embeds, but are outside of G, are by assumption *not* physically realizable in a world governed by the GPT G. The embedding relation must therefore be understood as a fundamental (rather than practical) limitation on what is possible in such a world.

[47] R. W. Spekkens, Evidence for the epistemic view of quantum states: A toy theory, Phys. Rev. A **75**, 032110 (2007).

[48] A. Cabello and G. García-Alcaine, Proposed Experimental Tests of the Bell-Kochen-Specker Theorem, Phys. Rev. Lett. **80**, 1797 (1998).

[49] A. Cabello, Kochen–Specker theorem and experimental test on hidden variables, Int. J. Mod. Phys. A **15**, 2813 (2000).

[50] C. Simon, M. Żukowski, H. Weinfurter, and A. Zeilinger, Feasible "Kochen-Specker" Experiment with Single Particles, Phys. Rev. Lett. **85**, 1783 (2000).

[51] C. Simon, i. c. v. Brukner, and A. Zeilinger, Hidden-Variable Theorems for Real Experiments, Phys. Rev. Lett. **86**, 4427 (2001).

[52] J.-Å.. Larsson, A Kochen-Specker inequality, Europhys. Lett. **58**, 799 (2002).

[53] A. Cabello, S. Filipp, H. Rauch, and Y. Hasegawa, Proposed Experiment for Testing Quantum Contextuality with Neutrons, Phys. Rev. Lett. **100**, 130404 (2008).

[54] S. Kochen and E. Specker, The problem of hidden variables in quantum mechanics, J. Math. Mech. 17, 59 (1967), also available from the Indiana Univ. Math. J.

[55] R. Kunjwal and R. W. Spekkens, From the Kochen-Specker Theorem to Noncontextuality Inequalities Without Assuming Determinism, Phys. Rev. Lett. **115**, 110403 (2015).

[56] R. Kunjwal, Contextuality beyond the Kochen-Specker theorem, arXiv:1612.07250 (2016).

[57] D. Schmid, J. Selby, and R. W. Spekkens, Unscrambling the omelette of causation and inference: The framework of causal-inferential theories, arXiv:2009.03297 (2020).

[58] P. Janotta and R. Lal, Generalized probabilistic theories without the no-restriction hypothesis, Phys. Rev. A **87**, 052131 (2013).

[59] P. Janotta and R. Lal, Non-locality in theories without the no-restriction hypothesis, arXiv:1412.8524 (2014).

[60] A. B. Sainz, Y. Guryanova, A. Acín, and M. Navascués, Almost-Quantum Correlations Violate the No-Restriction Hypothesis, Phys. Rev. Lett. **120**, 200402 (2018).

[61] S. N. Filippov, S. Gudder, T. Heinosaari, and L. Leppäjärvi, Operational restrictions in general probabilistic theories, Found. Phys. **50**, 850 (2020).

[62] V. J. Wright and S. Weigert, General Probabilistic Theories with a Gleason-type Theorem, arXiv:2005.14166 (2020).

[63] R. Kunjwal and R. W. Spekkens, From statistical proofs of the Kochen-Specker theorem to noise-robust noncontextuality inequalities, arXiv:1708.04793 (2017).

[64] R. Kunjwal, Beyond the Cabello-Severini-Winter framework: Making sense of contextuality without sharpness of measurements, Quantum **3**, 184 (2019).

[65] R. Kunjwal, Hypergraph framework for irreducible noncontextuality inequalities from logical proofs of the Kochen-Specker theorem, Quantum **4**, 219 (2020).

[66] M. F. Pusey, The robust noncontextuality inequalities in the simplest scenario, arXiv:1506.04178 (2015).