The Characterization of Noncontextuality in the Framework of Generalized Probabilistic Theories

David Schmid, John H. Selby, Elie Wolfe, Ravi Kunjwal, and Robert W. Spekkens
Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo, Ontario Canada N2L 2Y5
ICTQT, University of Gdańsk, Wita Stwosza 63, 80-308 Gdańsk, Poland

One of the most well-motivated and widely applicable notions of classicality for an operational theory is explainability 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-and-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 \textit{simplex-embeddability}. A simplex-embeddable GPT need not be classical in the traditional sense of its state space being a simplex and its effect space being the dual of this simplex; rather, it must admit of an explanation in terms of such a simplicial GPT. As such, simplex-embeddability constitutes an intuitive and freestanding notion of classicality for GPTs. Our result also provides a novel means of testing whether a given set of experimental data admits of such an explanation.

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 \cite{1, 2} and the failure to admit of a generalized-noncontextual ontological model \cite{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. 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 \cite{13} 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 quasi-probability representation \cite{14, 15} or of a locally causal model \cite{1, 2}. Additionally, the failure of noncontextuality has been shown to be a resource for quantum information processing \cite{16–24}. Finally, if one assumes the framework of ontological models, then assuming generalized noncontextuality can be understood as assuming an instance of a methodological principle for theory construction due to Leibniz (see Ref. \cite{25} and the appendix of Ref. \cite{12}). Given that Einstein made significant use of this principle when he developed the Theory of Relativity \cite{25}, it is seen to have impressive credentials in physics and therefore is a very 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.

It is our purpose here to understand what the existence of a generalized-noncontextual ontological model of an operational theory implies about the geometry of the generalized probabilistic theory \cite{26, 27} (GPT) which characterizes that operational theory. Ultimately we will prove the following:

\textbf{Theorem 1.} A prepare-and-measure operational scenario admits of a generalized-noncontextual ontological model on an ontic state space of finite cardinality if and only if the GPT which represents it is simplex-embeddable.

Simplex embeddability will be defined rigorously in Definition 1. It 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.

This theorem implies that 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 that a GPT should be deemed classical if and only if its state space is a simplex and its effect space the dual thereof: simplex-embeddability deems a GPT as classical if and only if it can be \textit{simulated} by such a simplex and its dual.
We begin by introducing operational theories and the generalized probabilistic theories associated to these, as well as ontological models of each.

**Operational Theories**—An *operational theory* is a very 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 $\text{Preps}$ and $\text{Mnts}$ respectively. These are conceived of as lists of lab instructions that one could, in principle, implement on the given system. We will 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 $\text{Effects}$, by considering the set of all outcomes $k$ for each measurement $M$ in the set $\text{Mnts}$. 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, however, is not completely arbitrary, but must be compatible with certain relations that hold between the procedures.\(^2\) For instance, if $P_1$ is described as a procedure that conveys mixtures $P_2$ 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 $[k|M]$ \[^{26}\].

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 $\Pr(\cdot, \cdot)$ 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 $T := \{\text{Preps}, \text{Mnts}, \Pr(\cdot, \cdot)\}$.

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 $[k|M] \in \text{Effects}$. 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 $P \in \text{Preps}$.

**Generalized Probabilistic Theories**—The full framework of GPTs allows for sequential and parallel composition; however, we will focus here 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, $\Omega$. 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 $\Omega$ is finite dimensional and compact. While $\Omega$ naturally lives inside an affine space, $\text{AffSpan}[\Omega]$, for convenience we will represent this as living inside a real inner product space $(V, (\cdot, \cdot))$ of one dimension higher, where we embed $\text{AffSpan}[\Omega]$ as a hyperplane in $V$ which 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.

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 $e \in \mathcal{E}$ given a state $s \in \Omega$ is given by the inner product:

$$\Pr(e, s) := \langle e, s \rangle. \quad (1)$$

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

$$\Omega^* := \{x \in V | (x, s) \in [0, 1] \forall s \in \Omega\}, \quad (2)$$

then $\mathcal{E}$ is a compact convex set contained in $\Omega^*$, $\mathcal{E} \subseteq \Omega^*$, which contains the origin $\mathbf{0}$ and the “unit effect” $\mathbf{u}$, which in turn satisfy, respectively, $(0, s) = 0$ and $(\mathbf{u}, s) = 1$ for all $s \in \Omega$. Due to how we embedded $\text{AffSpan}[\Omega]$ within $V$, $\mathbf{u}$ necessarily exists and is unique.\(^3\)

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 that $\langle e, s_1 \rangle = \langle e, s_2 \rangle$ for all $e \in \mathcal{E}$ if and only if $s_1 = s_2$, and for the GPT effects we have that $\langle e_1, s \rangle = \langle e_2, s \rangle$ for all $s \in \Omega$ if and only if $e_1 = e_2$.

A GPT, $G$, is therefore defined by $G := (V, (\cdot, \cdot), \Omega, \mathcal{E})$.

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

$$s : \text{Preps} \to \Omega \quad \text{and} \quad e : \text{Effects} \to \mathcal{E} \quad (3)$$

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

$$\Pr([k|M], P) = \langle e_{[k|M]}, s_P \rangle \quad (4)$$

for all preparations and effects in the operational theory.

---

\(^2\) In previous work and in the current article, these relations are *only* stipulated in the “ordinary language” description of the procedures, but in forthcoming work \[^{28}\] it is shown how they can be incorporated into the formal structure of an operational theory.

\(^3\) This uniqueness is referred to as the “causality axiom” in Ref. \[^{29}\].
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:

\[ P \simeq P' \iff s_P = s_{P'} \quad \text{and} \quad [k|M] \simeq [k'|M'] \quad \text{(5)} \]

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 will follow that \( s_{P_1} = w s_{P_2} + (1-w) s_{P_3} \).

### Ontological Model of an Operational Theory

An ontological model of an operational theory \( T \) is an attempt to explain the operational statistics of \( T \) by positing that the system has an ontic state, drawn from a set of ontic states \( \Lambda \), which determines (perhaps only probabilistically) the measurement outcome, and where there may be epistemic uncertainty about the ontic state for a given preparation procedure. Thus, it associates to each preparation \( P \in \text{Preps} \) a normalized probability distribution over \( \Lambda \), denoted \( \mu_P \), representing an agent’s knowledge of the ontic state when they know that the preparation was \( P \). Denoting the set of such distributions by \( D[\Lambda] \), the ontological model specifies a map

\[ \mu_L : \text{Preps} \to D[\Lambda]. \quad \text{(7)} \]

Furthermore, an ontological model associates to each operational effect \([k|M] \in \text{Effects}\) a response function on \( \Lambda \), denoted \( \xi_{[k|M]} \), where \( \xi_{[k|M]}(\lambda) \), represents the probability of obtaining the outcome \( k \) in a measurement of \( M \) when the ontic state of the system fed into the measurement device is \( \lambda \in \Lambda \). Denoting the set of such response functions by \( E[\Lambda] \), the ontological model specifies a map

\[ \xi_E : \text{Effects} \to E[\Lambda]. \quad \text{(8)} \]

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 \( \mu_{P_1} = w \mu_{P_2} + (1-w) \mu_{P_3} \), and similarly for operational effects. Finally, the ontological model must reproduce the probability rule of the operational theory \( T \) via

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

(where we have assumed \( \Lambda \) 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-and-measure experiment satisfies generalized noncontextuality if every two procedures which are operationally equivalent have identical representations in the ontological model.

In other words, the constraint for preparations is that

\[ P \simeq P' \implies \mu_P = \mu_{P'}. \quad \text{(10)} \]

while the constraint for operational effects is that

\[ [k|M] \simeq [k'|M'] \implies \xi_{[k|M]} = \xi_{[k'|M']}. \quad \text{(11)} \]

### Ontological Model of a GPT

Just like an ontological model of an operational theory (discussed above), an ontological model of a GPT is also an attempt to explain the operational statistics in terms of a space \( \Lambda \) of ontic states that the system can take and epistemic uncertainty thereof. In this case, however, what is being modelled ontologically are not preparations and measurements, but operational equivalence classes thereof. Thus, an ontological model of a GPT associates to each GPT state vector \( s \in \Omega \) a normalized probability distribution over \( \Lambda \), denoted \( \tilde{\mu}_s \in D[\Lambda] \), and to each GPT effect vector \( e \in E \) a response function on \( \Lambda \), denoted \( \tilde{\xi}_e \in E[\Lambda] \). Hence, it specifies a pair of maps

\[ \tilde{\mu}_L : \Omega \to D[\Lambda] \quad \text{and} \quad \tilde{\xi}_E : E \to E[\Lambda], \quad \text{(12)} \]

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

\[ \langle e, s \rangle = \sum_{\lambda \in \Lambda} \tilde{\xi}_e(\lambda) \tilde{\mu}_s(\lambda). \quad \text{(13)} \]

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 the appendix.) Hence:

**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 Ref. [14, 15] 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 when there is no variability of context for GPT states (or GPT effects) is a category mistake, in the same way that asking whether X varies with Y when Y exhibits no variability is a category mistake [30].

This implies another point of 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 (see, e.g., Refs. [31, 32]). 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.

**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. It is desirable to have this criterion expressed in the native language of GPTs, that is, in terms of convex-geometry. We will now give such a geometric 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 show that this restriction is irrelevant for many practical purposes.)

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

\begin{align}
  i(\Omega) &\subseteq \Delta_d, \\
  \kappa(\mathcal{E}) &\subseteq \Delta_d^*, \text{ and} \\
  \langle e, s \rangle_V = \langle \kappa(e), i(s) \rangle_W &\quad \forall e \in \mathcal{E}, \; s \in \Omega.
\end{align}

Note that while it is only the space of GPT *states* which 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 (as we show in the appendix).

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

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 which admit of generalized-noncontextual ontological models over ontic state spaces of finite cardinality.

**Discussion**— In the introduction, we listed several motivations for taking generalized noncontextuality as one’s notion of classicality for operational theories. By Theorem 1, each of these motivations can be reappropriated as a motivation for taking simplex-embeddability as one’s notion of classicality for a GPT. This notion of classicality is distinct from the prevailing notion that a GPT should be deemed classical if and only if its state space is a simplex and its effect space the dual of this simplex. Such a *simplicial GPT* is one for which all logically possible measurements can be performed simultaneously and without disturbance, and hence any such GPT is well-characterized as classical. However, a much broader class of GPTs is also well-characterized as classical: namely, all those which can be *simulated* by a simplicial GPT. This is exactly what our notion of simplex-embeddability captures.

The notion of simplex-embeddability is a more permissive notion of classicality than that of simpliciality since the class of simplex-embeddable GPTs is strictly larger than the class of simplicial GPTs. In related work undertaken simultaneously by Shahandeh [33] and by Barnum and Lami [34], it was shown that these two classes coincide if one restricts to GPTs which satisfy the no-restriction hypothesis [29]. Consequently, for these GPTs (and only these) the two notions of classicality coincide.

**Applications**— Our result provides a novel way to witness the nonclassicality (or classicality) of a given set of prepare-measure experiments on a system. One first infers the set of GPTs that are consistent with one’s data, using the techniques described in Ref. [35], and then one tests these GPTs for simplex-embeddability. This provides a very even-handed test of noncontextuality insofar as it does not privilege any particular preparations or measurements, but rather appeals to the geometry of the entire state space and the entire effect space.

We now address two important practical issues with testing for simplex-embeddability in such a scheme. Firstly, one needs some prescription for *how to test* whether a generic GPT $G$ admits an embedding of the type described in Definition 1 for a given dimension $d$. Secondly, even supposing that such an algorithm is available and efficient, a further obstacle is the lack of any *upper bound* on the dimension $d$ up to which one must apply this test. Because Theorem 2 does not, a priori, provide any bounds on $d$, it may be that one must repeat the embeddability test for a sequence of ever-larger dimensions $d$, and there is no guarantee that the process will terminate at any finite $d$.

A critical question, then, is whether there exists any finite dimension $b(G)$ such that if a GPT $G$ is simplex-embeddable, then that embedding can be constructed in a vector space of dimension $d \leq b(G)$. Without such a bound, one also cannot answer the question of whether...
there exist GPTs which only admit of ontological models where the cardinality of ontic states is uncountable.

In the special case where the GPT in question has an effect space that is a polytope, such a bound can be computed by leveraging arguments from Ref. [36]. Such a GPT is necessarily consistent with some operational theory defined by a finite set of operational effects. But the arguments in Ref. [36] imply that any prepare-and-measure scenario where the set of operational effects is of finite cardinality admits of a generalized-noncontextual ontological model if and only if it admits of one with at most a particular finite number (which we denote $d_{\text{max}}$) of ontic states. By the construction used to prove Proposition 1, this generalized-noncontextual ontological model for the operational theory can be translated into an ontological model for the GPT on the same finite-cardinality ontic state space.

To witness nonclassicality of a generic GPT, it suffices to find inner approximations of its state and effect spaces which do not embed in a simplex and its dual. Similarly, to witness classicality, it suffices to find outer approximations of these that do embed in a simplex and its dual. In practice, one can always take the (inner or outer) approximations of the effect space to be polytopic, in which case the dimension bound described above can be leveraged. Note further that in a real experiment, the data will not pick out a single GPT but a set of compatible possibilities [35]; to witness nonclassicality (respectively classicality) of the GPT that generated the experimental data, then, it suffices to find a polytopic inner (respectively outer) approximation of all the GPTs in this set.

**Outlook**— A key question for future work is to extend our result beyond the prepare-and-measure scenario to more general compositional scenarios. As discussed above, it would also be interesting to remove the dimensional caveats from our main theorems, and to find efficient methods for testing whether a given GPT is simplex-embeddable or not.

**ACKNOWLEDGMENTS**

We thank Matt Pusey and Thomas Galley for helpful discussions. RWS 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. JHS was supported in part by the Foundation for Polish Science through IRAP project co-financed by EU within Smart Growth Operational Programme (contract no. 2018/MAB/5). This research was supported by Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is supported by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Research, Innovation and Science.

---

4 Translating the results of Ref. [36] into the language of GPTs, it is simply the number of convexly extremal states in the set of logically possible states, i.e., the number of vertices of $L^*$.
random access codes," New J. Phys. 18, 045003 (2016).
[18] Andris Ambainis, Manik Banik, Anubhav Chaturvedi, Dmitry Kravchenko, and Ashutosh Rai, “Parity Oblivious d-Level Random Access Codes and Class of Noncontextuality Inequalities," arXiv:1607.05490 (2016).
[19] Debashis Saha, Pawel Horodecki, and Marcin Pawlowski, “State independent contextuality advances one-way communication," New J. Phys. 21, 093057 (2019).
[20] David Schmid and Robert W. Spekkens, “Contextual Advantage for State Discrimination," Phys. Rev. X 8, 011015 (2018).
[21] Matteo Lostaglio and Gabriel Senno, “Contextual advantage for state-dependent cloning," arXiv:1905.08291 (2019).
[22] Mark Howard, Joel Wallman, Victor Veitch, and Joseph Emerson, “Contextuality supplies the ‘magic’ for quantum computation," Nature 510, 351–355 (2014).
[23] Robert Raussendorf, “Contextuality in measurement-based quantum computation," Phys. Rev. A 88, 022322 (2013).
[24] Matty J Hoban, Earl T Campbell, Klearchos Loukopoulos, and Dan E Browne, “Non-adaptive measurement-based quantum computation and multi-party Bell inequalities," New J. Phys. 13, 023014 (2011).
[25] Robert W. Spekkens, “The ontological identity of empirical indiscernibles: Leibniz’s methodological principle and its significance in the work of Einstein," arXiv:1909.04628 (2019).
[26] Lucien Hardy, “Quantum Theory From Five Reasonable Axioms," quant-ph/0101012 (2001).
[27] Jonathan Barrett, “Information processing in generalized probabilistic theories," Phys. Rev. A 75, 032304 (2007).
[28] D. Schmid, J. Selby, and R. W. Spekkens, “Unscrambling the causal-inferential omelette in operational and ontological theories," Forthcoming.
[29] Giulio Chiribella, Giacomo Mauro D’Ariano, and Paolo Perinotti, “Probabilistic theories with purification," Phys. Rev. A 81, 062348 (2010).
[30] Ofra Magidor, “Category Mistakes," in The Stanford Encyclopedia of Philosophy (Metaphysics Research Lab, Stanford University, 2019).
[31] David Bohm, “A Suggested Interpretation of the Quantum Theory in Terms of ‘Hidden’ Variables. I," Phys. Rev. 85, 166–179 (1952).
[32] E G Beltranmetti and S Bugajski, “A classical extension of quantum mechanics," J. Phys. A 28, 3329–3343 (1995).
[33] Farid Shahandeh, “Contextuality of general probabilistic theories and the power of a single resource," Forthcoming.
[34] Howard Barnum and Ludovico Lami, “Classical embeddability of general probabilistic theories," Forthcoming.
[35] Michael D Mazurek, Matthew F Pusey, Kevin J Resch, and Robert W Spekkens, “Experimentally bounding deviations from quantum theory in the landscape of generalized probabilistic theories," arXiv:1710.05948 (2017).
[36] David Schmid, Robert W. Spekkens, and Elie 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).
[37] Robert W. Spekkens, “Evidence for the epistemic view of quantum states: A toy theory," Phys. Rev. A 75, 032110 (2007).

Appendix A: Proof of Proposition 1

We now prove Proposition 1.

Proof. In one direction, given an ontological model of the GPT G (i.e. given the relevant maps \( \tilde{\mu}_z \) and \( \tilde{\xi}_e \)) which is associated to the operational theory \( T \) (via the quotienting maps \( s_\lambda \) and \( e_\lambda \)) 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

\[ \mu_z := \tilde{\mu}_{s_\lambda} \]  

(A1)

and

\[ \xi_e := \tilde{\xi}_{e_\lambda}. \]  

(A2)

It is then easy to check that these maps define an ontological model of the operational theory \( T \). Firstly, \( \mu_z \) preserves the convex relations among preparations, since \( s_\lambda \) preserves these relations and \( \tilde{\mu}_z \) is linear. Similarly, \( \xi_e \) preserves the convex and coarse-graining relations for operational effects. The resulting model is easily seen to be noncontextual, since

\[ P_1 \otimes P_2 \Rightarrow s_{P_1} = s_{P_2} \]  

(A3)

\[ \Rightarrow \tilde{\mu}_{s_{P_1}} = \tilde{\mu}_{s_{P_2}} \]  

(A4)

\[ \Rightarrow \mu_{P_1} = \mu_{P_2} \]  

(A5)

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

\[ \sum_{\lambda \in \Lambda} \xi_{|k,M|}(\lambda) \mu_P(\lambda) = \sum_{\lambda \in \Lambda} \tilde{\xi}_{e_{|k,M|}}(\lambda) \tilde{\mu}_{s_P}(\lambda) \]  

(A6)

\[ \frac{1}{13} \leq \Pr(|k,M|; P) \]  

(A7)

\[ \frac{1}{13} = \Pr(|k,M|; P). \]  

(A8)

Conversely, given a generalized-noncontextual ontological model of an operational theory \( T \) (i.e. given the relevant maps \( \mu_z \) and \( \xi_e \)) we can construct an ontological model of the GPT \( G \) associated to it (via the relevant maps \( s_\lambda \) and \( e_\lambda \)) by defining the maps \( \tilde{\mu}_z \) and \( \tilde{\xi}_e \) as the unique linear maps satisfying

\[ \tilde{\mu}_{s} := \mu_{P} \]  

(A9)

for all preparations \( P \) such that \( s_P = s \), and

\[ \tilde{\xi}_{e} := \xi_{e_{|k,M|}} \]  

(A10)

for all operational effects \( |k,M| \) such that \( e_{|k,M|} = e \). It is then easy to check that \( \tilde{\mu}_z \) preserves the convex relations between state vectors. E.g., suppose that \( P_1 \) is a convex mixture of \( P_2 \) and \( P_3 \) with weights \( w \) and \( 1-w \), and hence \( \mu_{P_1} = w \mu_{P_2} + (1-w) \mu_{P_3} \) by fact that \( \mu_z \) preserves the convex relations among preparation procedures. We therefore find that \( \tilde{\mu}_{s_{P_1}} = \mu_{P_1} = w \mu_{P_2} + (1-w) \mu_{P_3} = w \tilde{\mu}_{s_{P_2}} + (1-w) \tilde{\mu}_{s_{P_3}} \). Hence whenever \( s_{P_1} = ws_{P_2} + (1-w)s_{P_3} \) (which
follows from the operational equivalence) holds, so does \( \mu_{\omega_{\pi_1}} = w \mu_{\pi_1} + (1 - w) \mu_{\pi_2} \), and so \( \mu \) does indeed preserve the convex relations among preparation procedures. The proof that \( \xi \) preserves the convex and coarse-graining relations among GPT effects is analogous.

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

\[
\sum_{\lambda \in \Lambda} \xi_{\text{e}}(\lambda) \mu_{\text{s}}(\lambda) = \sum_{\lambda \in \Lambda} \xi_{[k|M]}(\lambda) \mu_{P}(\lambda)
\]  

( Pistons 11)

\[
\frac{9}{1} \Pr(P_{[k|M]})(\lambda)
\]  

( Pistons 12)

\[
\frac{4}{1} \langle e_{[k|M]}, s_P \rangle
\]  

( Pistons 13)

\[
\langle e, s \rangle.
\]  

( Pistons 14)

☐

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 of \( G \).

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

\[
\hat{\mu} : \Omega \rightarrow \mathbb{R}^{\Lambda}.
\]  

(B1)

It also associates to each GPT effect \( e \) a real-valued function over \( \Lambda \), denoted \( \xi_e \) and satisfying \( \sum_{\lambda \in \Lambda} \xi_e(\lambda) = 1 \) for all \( \lambda \in \Lambda \) where \( \chi \) is any set of GPT effects corresponding to the set of outcomes of a possible measurement. Note that such a set must satisfy \( \sum_{e \in \chi} e = u \). The function \( \xi_e \) 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

\[
\xi : \mathcal{E} \rightarrow \mathbb{R}^{\Lambda}.
\]  

(B2)

Now consider the probability of obtaining GPT effect \( e \) given a preparation associated to state \( s \), which is given by \( \langle e, s \rangle \) in the GPT. In a quasiprobability representation of this GPT, one computes this probability using the same formula that would be appropriate if the quasiprobability were true probabilities, namely,

\[
\langle e, s \rangle_{V} = \xi_{e} \cdot \hat{\mu}_{s} = \sum_{\lambda \in \Lambda} \xi_{e}(\lambda) \hat{\mu}_{s}(\lambda).
\]  

(B3)

Finally, we will say that a quasiprobability representation of a GPT is positive if \( \forall s \in \Omega, \forall \lambda \in \Lambda: 0 \leq \hat{\mu}_{s}(\lambda) \leq 1 \) and \( \forall e \in \mathcal{E}, \forall \lambda \in \Lambda: 0 \leq \xi_{e}(\lambda) \leq 1 \). Prop 2 follows immediately from comparison of Eqs. (B1), (B2) and (B3) with Eqs. (12) and (13).

Appendix C: Proof of Theorem 2

We now wish to prove Theorem 2.

It is convenient for the proof to formally define the notion of a *simplicial GPT* which was discussed in the main text. A simplicial GPT of given dimension \( d \) is a tuple \((V, \langle \cdot, \cdot \rangle_{V}, \Omega, \mathcal{E})\) defined by a vector space \( V \) with inner product \( \langle \cdot, \cdot \rangle_{V} \), a simplicial state space \( \Omega = \Delta_{d} \) (of intrinsic dimension \( d \)), and the dual hypercube \( \tilde{\mathcal{E}} = \Delta_{d}^{\ast} \) (of intrinsic dimension \( d \)) as its effect space. Clearly, the inner product space, simplex, and the dual appearing in the definition of simplex embeddability (Def. 1) define such a simplicial GPT.\(^5\)

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 are essentially equivalent.

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

\[
\omega(\Omega) = \Omega',
\]  

(C1)

\[
\epsilon(\mathcal{E}) = \mathcal{E}', \quad \text{and}
\]  

(C2)

\[
\langle e, s \rangle_{V} = \langle \epsilon(e), \omega(s) \rangle_{V'}, \quad \forall e \in \mathcal{E}, \ s \in \Omega.
\]  

(C3)

As a simple example, note that \( (V, \langle \cdot, \cdot \rangle_{V}, \Omega, T(\mathcal{E})) \) is equivalent to \( (V, \langle \cdot, \cdot \rangle_{V}, T(\Omega), \mathcal{E}) \) where \( T \) is any reversible map between vector spaces.

Now, note that any simplicial GPT of given dimension \( d \) is equivalent (in the above sense) to one in a particular canonical form, defined as follows. The vector space is

\[
\hat{\mu} : \Omega \rightarrow \mathbb{R}^{\Lambda}.
\]  

(B1)

\[
\xi : \mathcal{E} \rightarrow \mathbb{R}^{\Lambda}.
\]  

(B2)

\[
\langle e, s \rangle_{V} = \xi_{e} \cdot \hat{\mu}_{s} = \sum_{\lambda \in \Lambda} \xi_{e}(\lambda) \hat{\mu}_{s}(\lambda).
\]  

(B3)

\[\text{for a GPT } G \text{ which is simplex-embeddable but not itself simplicial, the embedding relation cannot be interpreted as merely due to technological limitations. To posit that a world is governed by the GPT } G \text{ is to posit that the states and effects that are physically realizable are all and only the states and effects included in } G. \text{ Consequently, any states and effects that are included in the simplicial GPT into which } G \text{ embeds, but are outside of } G, \text{ are by assumption not physically realizable. The embedding relation must therefore be understood as a fundamental (rather than practical) limitation on what is possible in a world governed by such a GPT.}\]
taken to be \( \mathbb{R}^d \) with some chosen orthonormal basis \( \{b_i\}_{i \in B} \) 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
\[
\Delta_d := \text{ConvexHull}(\{b_i\}), \tag{C4}
\]
and the effect space is the dual of it, namely the unit hypercube (with \( 2^{d-1} \) vertices and intrinsic dimension \( d \))
\[
\Delta_d^* := \text{ConvexHull}(\{\sum_{i \in \chi} b_i | \chi \subseteq B \}). \tag{C5}
\]

Finally, the unit effect is \( u = \sum_{i \in B} b_i \).

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 model.

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

Next, one can identify the GPT vector space \( V \) with the vector space \( \mathbb{R}^\Lambda \) of functions from \( \Lambda \) 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 \( w \in V \) is uniquely identified with a function \( f_w \in \mathbb{R}^{\text{Vert}[\Delta_d]} \) by \( f_w(i) = \langle w, b_i \rangle_V \), where \( b_i \) is the orthonormal basis used in the definition of the canonical form (it is the set of vectors that point at the vertices of \( \Delta_d \)). Clearly, then, the vertices of \( \Delta_d \) will be represented by point distributions in \( \mathbb{R}^{\text{Vert}[\Delta_d]} \), as \( f_{b_i}(i) = \langle b_i, b_i \rangle_V = \delta_{ij} \). By linearity, the other vectors in \( \Delta_d \) will correspond to mixtures of these, and hence to probability distributions over \( \text{Vert}[\Delta_d] \). As such, there is a one-to-one correspondence \( D[\Lambda] \leftrightarrow \Delta_d \).

Similarly, the vectors in \( \Delta_d^* \) can be identified with the response functions. The extremal points of \( \Delta_d^* \), namely, \( \sum_{i \in \chi} b_i \), are represented by the functions: \( f_{\sum_{i \in \chi} b_i}(i) = \sum_{i \in \chi} \langle b_i, b_i \rangle_V = \sum_{i \in \chi} \delta_{ij} \), and these correspond to the extremal response functions in \( F[\Lambda] \). Again by linearity, we see that every vector in \( \Delta_d^* \) is identified with a valid response function. Additionally, the vector \( u \) 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 \( \mathbf{1} \leftrightarrow u \).

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

We can summarize all this in a table:

| ontological | geometric |
|------------|-----------|
| \( \lambda \in \Lambda \) | \( i \in \text{Vert}[\Delta_d] \) |
| \( \mathbb{R}^\Lambda \) | \( V \) |
| \( D[\Lambda] \) | \( \Delta_d \) |
| \( F[\Lambda] \) | \( \Delta_d^* \) |
| \( \mathbf{1} \) | \( u \) |
| \( \sum_{\lambda \in \Lambda} f_w(\lambda) f_{w'}(\lambda) \) | \( \langle w, w' \rangle_V \) |

Hence, any simplex-embeddable GPT can be embedded into a simplicial GPT, whose canonical form together with the identification in the table can be used to explicitly construct an ontological representation of that GPT.

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

This establishes the theorem.

**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.

The GPT state space of the stabilizer rebit is the convex subset
\[
\text{ConvexHull}(\{\langle 0|0\rangle, \langle 1|1\rangle, \langle +|+\rangle, \langle -|-\rangle \}) \tag{D1}
\]
of the full qubit state space:

\[
\begin{array}{c|cc|c|c}
\hline |
\hline
| \begin{array}{c}
0\rangle \\langle 0|
\end{array} & | \begin{array}{c}
1\rangle \\langle 1|
\end{array} & | \begin{array}{c}
+\rangle \\langle +|
\end{array} & | \begin{array}{c}
-\rangle \\langle -|
\end{array} \\
| \hline
\end{array}
\]

and the effect space is the convex subset
\[
\text{ConvexHull}(\{0,0\rangle \langle 0|, 1, 1\rangle \langle 1|, +, +\rangle \langle +|, -, -\rangle \langle -|, 1 \}) \tag{D2}
\]
of the full qubit effect space.

This stabilizer rebit GPT admits of an ontological model; one such model is the toy model of Ref. [37]. 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. [37] defines a tetrahedron which contains the state space, and whose dual contains the effect space. Note that the stabilizer rebit state space is 2-dimensional, while the simplex (the tetrahedron) in which it is embedded is intrinsically 3-dimensional. As we now prove, this dimension mismatch is unavoidable: it is not possible to find an embedding into any intrinsically 2-dimensional simplex such that the effect space embeds in the dual.

**Proof.** We assume that there exists a triangle which 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 $\text{tr}[[1\rangle\langle 1]|\rho] = \langle 1|\rho|1\rangle = 0$, can be geometrically represented by a hyperplane in the state space:

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 which 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, that is from those of the GPT. Hence, $|0\rangle\langle 0|$ and $|+\rangle\langle +|$ must be mapped to disjoint faces, for example:

We can apply this same reasoning to each of the four extremal states to conclude that they must be mapped to four disjoint faces of the triangle. But there are at most three disjoint faces of a triangle, so we reach a contradiction.

### a. 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. [37]. The ontic state space is defined as

$$\Lambda \equiv \{\lambda_0, \lambda_0, \lambda_1, \lambda_1\}. \quad (D3)$$

The epistemic states are given by

$$\bar{\mu}_0 = (1/2,1/2,0,0), \quad \bar{\mu}_1 = (0,0,1/2,1/2), \quad \bar{\mu}_+ = (1/2,0,1/2,0), \quad \bar{\mu}_- = (0,1/2,0,1/2). \quad (D4)$$

The response functions are given by

$$\xi_0 = (1,1,0,0), \quad \xi_+ = (1,0,1,0), \quad \xi_1 = (0,0,1,1), \quad \xi_- = (0,1,0,1). \quad (D5)$$

It is easy to see that this model reproduces the operational predictions, since

$$p(0|M_{01},P_0) = \bar{\xi}_0 \cdot \bar{\mu}_0 = 1,$$

$$p(0|M_{01},P_1) = \bar{\xi}_0 \cdot \bar{\mu}_1 = 0,$$

$$p(0|M_{01},P_+) = \bar{\xi}_0 \cdot \bar{\mu}_+ = 1/2,$$

$$p(0|M_{01},P_-) = \bar{\xi}_0 \cdot \bar{\mu}_- = 1/2,$$

$$p(+)|M_+,P_0) = \bar{\xi}_+ \cdot \bar{\mu}_0 = 1/2,$$

$$p(+)|M_+,P_1) = \bar{\xi}_+ \cdot \bar{\mu}_1 = 1/2,$$

$$p(+)|M_+,P_+) = \bar{\xi}_+ \cdot \bar{\mu}_+ = 1,$$

$$p(+)|M_+,P_-) = \bar{\xi}_+ \cdot \bar{\mu}_- = 0. \quad (D6)$$