Linear Algebra of Generalized Contextuality in Prepare-Transform-Measure Scenarios

Theodoros Yianni Nyan Raess Farid Shahandeh
Department of Computer Science
Royal Holloway, University of London
Egham, United Kingdom
theo.yianni.2023@live.rhul.ac.uk
Nyan.Raess.2024@live.rhul.ac.uk
farid.shahandeh@rhul.ac.uk

1 Introduction

Quantum contextuality is the impossibility of reproducing the statistical predictions of the theory with an ontological model in which operationally indistinguishable procedures are represented identically [1, 2]. In the current work, we refer to the notion of *generalized contextuality*, first introduced in Ref. [1]. This property is known to underlie quantum advantage in various information-processing tasks, encompassing quantum computations [3–5] and quantum communications [6, 7].

Following from the work on prepare-measure scenarios in Ref. [8], we develop a statistics-first, linear-algebraic framework to test generalized contextuality in operational scenarios with transformations beyond the prepare–measure setting. First, we extend the rank-based characterization from prepare–measure scenarios to prepare–transform–measure scenarios by replacing the COPE matrix with an analogous COPE tensor. We then introduce higher-order transformation scenarios, where operations act on sets of lower-level transformations, and give a full decision procedure for contextuality together with a complexity analysis. Using this higher-order description, we derive a contextuality condition for sequential scenarios with multiple transformation stages. The advantage of this is that it can be checked with a similar complexity to deciding the contextuality of prepare-measure scenarios. We construct a toy theory in which contextuality can be detected by our condition, but vanishes when the sequential structure of transformations is collapsed.

2 PTM Scenarios

The simplest scenario we consider is the prepare-transform-measure scenario with a single transformation stage, denoted by $\mathsf{PT}^1\mathsf{M}$. We can organize the conditional outcome probabilities of the scenario into a tensor C, the entries of which are the probabilities $p(k|i,j,l) := p(k|\mathsf{P}_i,\mathsf{T}^1_j,\mathsf{M}_l)$ of each outcome E^l_k, given that the system was prepared, transformed and measured according to $\mathsf{P}_i, \mathsf{T}^1_j$ and M_l respectively. In a $\mathsf{PT}^1\mathsf{M}$ scenario, C is therefore a 3-tensor of *conditional outcome probabilities of events* (COPE). It is assumed that C represents the entire probabilistic structure of the operational theory admissible for such scenarios [8].

An ontological model of the $\mathsf{PT}^1\mathsf{M}$ scenario is given by a nonnegative tensor factorization of its COPE tensor C of the form,

$$C_{ijk} = R_{i:}\Gamma^{1\,j}E_{:k} = p(l|\mathsf{P}_k,\mathsf{T}^1_j), \tag{1}$$

Submitted to:
QPL 2026

where we have re-indexed the outcomes using a single global index, rather than a pair consisting of the measurement label and the measurement outcome label. Here, the rows of R are the response functions of the ontological model representing measurement outcomes, the columns of E are the epistemic states representing preparations, and the stochastic matrix $\Gamma^{1\,j}$ represents the jth transformation from the set of available transformation procedures.

Our first main result is the following.

Theorem 1. *The operational theory of the* $\mathsf{PT}^1\mathsf{M}$ *scenario admits a noncontextual ontological model (NCOM) iff its COPE 3-tensor admits a nonnegative tensor factorization* $C_{ijk} = R_{i:}\Gamma^{1\,j}E_{:k}$ *such that*

$$\begin{aligned} \operatorname{rank}(R) &= \operatorname{rank}(C_{[\mathsf{E}]}), \\ \operatorname{rank}(E) &= \operatorname{rank}(C_{[\mathsf{P}]}), \\ \operatorname{rank}(G^1) &= \operatorname{rank}(C_{[\mathsf{T}^1]}). \end{aligned} \tag{2}$$

Here, $C_{[\mathsf{E}]}$, $C_{[\mathsf{T}^1]}$, *and* $C_{[\mathsf{P}]}$ *denote the mode-*E, *mode-*T^1 *and mode-*P *flattenings [9] of the COPE tensor* C, *respectively. Furthermore,* G^1 *represents the matrix of row-flattened* $\Gamma^{1\,j}$ *matrices.*

The proof of this theorem is given in the full paper.

3 Computing Noncontextual Models of $\mathsf{PT}^1\mathsf{M}$ Scenarios

To formulate the contextuality of $\mathsf{PT}^1\mathsf{M}$ scenarios linear algebraically, we first derive the following lemma, which will be used in constructing noncontextual ontological models (NCOMs).

Lemma 1. *Suppose a* $\mathsf{PT}^1\mathsf{M}$ *scenario with COPE tensor* C *admits NCOMs of the form* $C_{ijk} = R_{i:}\Gamma^{1\,j}E_{:k}$. *Then there exist matrices* $\bar{R}$ *and* $\bar{E}$, *depending only on* C, *called* extremal factors, *such that* $\operatorname{rank}(\bar{R}) = \operatorname{rank}(C_{[\mathsf{E}]})$ *and* $\operatorname{rank}(\bar{E}) = \operatorname{rank}(C_{[\mathsf{P}]})$. *Moreover, for all NCOMs, the corresponding factors* R *and* E *can be written as* $R = \bar{R}M$ *and* $E = M'\bar{E}$ *for some nonnegative* M, M'.

Due to space limitations, we defer the proof to the supplementary paper. This leads to the following lemma.

Lemma 2. *It can be decided whether an operational theory with* n_{P} *preparations,* n_{E} *measurement events,* n_{T^1} *transformations, and the greatest Tucker rank component* r *of its COPE tensor admits a noncontextual ontological model in time poly*$((n_{\mathsf{E}}n_{\mathsf{P}})^{O(r)}, n_{\mathsf{T}^1})$.

Proof. (outline). We first observe that, by Lemma 1, for an ontological model given by response function matrix R, epistemic state matrix E, and stochastic transformation matrices $\{\Gamma^{1\,i}\}_{i=1}^{n_{\mathsf{T}^1}}$, the slice of the COPE tensor given by $R\Gamma^{1\,i}E$ can always be written in terms of extremal factors as $\bar{R}(M\Gamma^{1\,i}M')\bar{E}$, for some nonnegative M and M'. Then, the matrices M and M' can be absorbed into the $\Gamma^{1\,i}$ matrices. It follows that if an NCOM exists, there exists one with $R = \bar{R}$ and $E = \bar{E}$. Therefore, only the representation of the transformations needs to be optimized, which can be done with a linear program. Given that $\bar{R}$ and $\bar{E}$ have $n_{\mathsf{E}}^{O(r)}$ columns and $n_{\mathsf{P}}^{O(r)}$ rows, respectively, as shown in the full paper, $(n_{\mathsf{E}}n_{\mathsf{P}})^{O(r)}n_{\mathsf{T}^1}$ variables need to be optimized, hence the complexity result. ■

4 Computing Noncontextual Models of $\mathsf{PT}^2[\mathsf{T}^1]\mathsf{M}$ Scenarios

We extend our formalism to operational theories with different causal structures, rather than just one stage of transformations, in particular, the nested $\mathsf{PT}^2[\mathsf{T}^1]\mathsf{M}$ scenario. Here, $\mathsf{T}^2[\cdot]$ is a second-order transformation, meaning that it acts on the system's transformations. An ontological model of the $\mathsf{PT}^2[\mathsf{T}^1]\mathsf{M}$

scenario is given by a nonnegative tensor factorization of its COPE tensor C of the form,

$$C_{ijkl} = \sum_{wxyz} R_{iw} \Gamma^{2k}_{wxyz} \Gamma^{1j}_{yz} E_{xl}. \tag{3}$$

Here, the 4-tensor Γ^{2k} represents the kth second-order transformation setting in the operational theory.

Define the matrix G^2 as the matrix of row-flattened Γ^{2k} tensors, and $C_{[\mathsf{T}^2]}$ as the mode-T^2 flattening of C. Then, the noncontextuality conditions for this scenario are given by Eq. (2) and the following:

$$\operatorname{rank}(G^2) = \operatorname{rank}(C_{[\mathsf{T}^2]}). \tag{4}$$

Theorem 2. *It can be decided whether an operational theory with n_{P} preparations, n_{E} measurement events, n_{T^1} first-order transformations, n_{T^2} second-order transformations and the greatest Tucker rank component r of its COPE tensor admits a noncontextual ontological model in time poly$((n_{\mathsf{E}} n_{\mathsf{P}} n_{\mathsf{T}^1})^{O(r)}, n_{\mathsf{T}^2})$.*

Proof. There exist extremal factors $\bar{R}$, $\bar{E}$ and $\{\bar{\Gamma}^{1j}\}$, depending only on C, satisfying the corresponding noncontextuality conditions. Moreover, for all NCOMs, the corresponding factors R, E and Γ^{1j} can be written as $R = \bar{R}M$, $E = M'\bar{E}$ and $\Gamma^{1j}_{wx} = \sum_{yz} M''_{wxyz} \bar{\Gamma}^{1j}_{yz}$, for some nonnegative M, M', M''. According to Eq. (3), R, E, and Γ^{1j} only contract with Γ^{2k}, therefore, M, M' and M'' can be absorbed into Γ^{2k}. Now, only the representation of T^2 needs to be optimized. The complexity result follows from a line of reasoning similar to that of the $\mathsf{PT}^1\mathsf{M}$ case. ■

This result can be extended to higher-order scenarios in the obvious way. For example, in the $\mathsf{P}(\mathsf{T}^3[\mathsf{T}^2])[\mathsf{T}^1]\mathsf{M}$ scenario, where T^3 is third-order, the complexity would become $poly((n_{\mathsf{E}} n_{\mathsf{P}} n_{\mathsf{T}^1} n_{\mathsf{T}^2})^{O(r)}, n_{\mathsf{T}^3})$.

5 Application to Sequential PTM Scenarios

The simplest example of a sequential PTM scenario is the $\mathsf{PT}^1\mathsf{T}^2\mathsf{M}$ scenario, where T^1 and T^2 are both first-order and act on the system sequentially. This can be viewed as a special case of $\mathsf{PT}^2[\mathsf{T}^1]\mathsf{M}$ scenarios. Using Theorem 2, we arrive at the following contextuality condition.

Theorem 3. *A $\mathsf{PT}^1\mathsf{T}^2\mathsf{M}$ scenario is contextual if interpreting the same COPE tensor as that of a $\mathsf{PT}^2[\mathsf{T}^1]\mathsf{M}$ scenario is contextual.*

Note that Theorem 3 provides only a sufficient, and not a necessary, condition for the contextuality of $\mathsf{PT}^1\mathsf{T}^2\mathsf{M}$ scenarios. We note also that this can be extended in the obvious way to more stages of sequential transformations.

6 An Example

In the corresponding paper of this abstract, we provide an example of a contextual $\mathsf{PT}^1\mathsf{T}^2\mathsf{M}$ scenario where contextuality is removed when the scenario is coarse-grained by lumping operation stages together. However, it emerges when the causal structure is enforced, and it can be certified using Theorem 3, Eq. (2), and Eq. (4). This shows that our condition detects contextuality missed by existing methods.

References

[1] R. W. Spekkens. Contextuality for preparations, transformations, and unsharp measurements. *Physical Review A*, 71(5), May 2005. ISSN 1094-1622. URL http://dx.doi.org/10.1103/PhysRevA.71.052108.

[2] Ravi Kunjwal and Robert W. Spekkens. From the kochen-specker theorem to noncontextuality inequalities without assuming determinism. *Phys. Rev. Lett.*, 115:110403, Sep 2015. doi: 10.1103/PhysRevLett.115.110403. URL https://link.aps.org/doi/10.1103/PhysRevLett.115.110403.

[3] Juan Bermejo-Vega, Nicolas Delfosse, Dan E. Browne, Cihan Okay, and Robert Raussendorf. Contextuality as a resource for models of quantum computation with qubits. *Phys. Rev. Lett.*, 119: 120505, Sep 2017. URL https://link.aps.org/doi/10.1103/PhysRevLett.119.120505.

[4] Farid Shahandeh. Quantum computational advantage implies contextuality, 2021. URL https://arxiv.org/abs/2112.00024.

[5] David Schmid, Haoxing Du, John H. Selby, and Matthew F. Pusey. Uniqueness of noncontextual models for stabilizer subtheories. *Phys. Rev. Lett.*, 129:120403, Sep 2022. URL https://link.aps.org/doi/10.1103/PhysRevLett.129.120403.

[6] Shashank Gupta, Debashis Saha, Zhen-Peng Xu, Adán Cabello, and A. S. Majumdar. Quantum contextuality provides communication complexity advantage. *Phys. Rev. Lett.*, 130:080802, Feb 2023. URL https://link.aps.org/doi/10.1103/PhysRevLett.130.080802.

[7] Debashis Saha and Anubhav Chaturvedi. Preparation contextuality as an essential feature underlying quantum communication advantage. *Phys. Rev. A*, 100:022108, Aug 2019. doi: 10.1103/PhysRevA.100.022108. URL https://link.aps.org/doi/10.1103/PhysRevA.100.022108.

[8] Farid Shahandeh, Theodoros Yianni, and Mina Doosti. A unified linear algebraic framework for physical models and generalized contextuality, 2025. URL https://arxiv.org/abs/2512.10000.

[9] *Introduction – Problem Statements and Models*, chapter 1, pages 1–80. John Wiley Sons, Ltd. ISBN 9780470747278. doi: https://doi.org/10.1002/9780470747278.ch1. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470747278.ch1.

Linear Algebra of Generalized Contextuality in Prepare-Transform-Measure Scenarios

Theodoros Yianni,[1] Nyan Raess,[1] and Farid Shahandeh[1]

[1] *Department of Computer Science, Royal Holloway, University of London, United Kingdom*

Generalized contextuality is a canonical distinguishing property of quantum theory. Existing methods for testing generalized contextuality of general probabilistic theories, in particular quantum mechanics, are well developed for prepare-measure and single-stage prepare-measure-transform scenarios. In a recent work [arXiv:2512.10000], a bottom-up, statistics-first linear-algebraic framework for contextuality in prepare-measure scenarios was introduced. We extend this approach to operational scenarios with sequential and higher-order transformations and give a full decision procedure for contextuality and analyze its computational complexity. We then show how the higher-order framework yields a one-sided test for multistage sequential prepare–transform–measure scenarios, providing a sufficient condition for noncontextuality that can be checked faster than a full sequential decision procedure. We construct an operational theory in which contextuality manifests itself only in the sequential structure of the transformations. Our results sheds new light on the compositional aspects of generalized contextuality.

I. Introduction

A central problem of quantum foundations is explaining the separation between classical and nonclassical theories. A canonical distinguishing property of the latter is contextuality, which is the impossibility of reproducing the statistical predictions of the theory with an ontological model in which operationally indistinguishable procedures are represented identically [1, 2]. Various widely used notions and frameworks of contextuality include Kochen-Specker contextuality [3], the sheaf-theoretic framework [4], Contextuality-by-Default [5], state-independent contextuality [6], and graph-theoretic formulations [7]. Also, Bell's work on nonlocality [8, 9] is closely related and can be viewed as contextuality in multipartite measurement scenarios [4, 10].

In the current work, we refer to the notion of *generalized contextuality*, first introduced in Ref. [1]. This property is known to underlie quantum advantage in various information-processing tasks, encompassing quantum computations [11–13] and quantum communications [14, 15].

Many information-processing tasks of interest are naturally described as sequential compositions of operations. In such settings, classical simulability and classicality are often not properties of any single step in isolation, but rather of how the steps compose. In such protocols, it may be that each individual stage admits a classical explanation, even though the overall process does not. Conversely, one can also devise scenarios in which the compositional process does not admit a noncontextual ontological explanation, even though such an explanation exists when all the transformations are treated as a single, combined operation. This motivates studying contextuality in scenarios with multiple transformation stages, where one can ask how contextuality behaves under concatenation and coarse-graining. In particular, computation is implemented by circuits whose power is determined by the composition of stages, making multistage contextuality a natural candidate property for explaining circuit-level quantum advantage [12, 16, 17].

Despite this, contextuality has been most extensively studied in the simplest possible scenario: the prepare-measure (PM) scenario, in which an experimenter has access to a finite number of ways to prepare and measure the system. Many tools already exist to probe the contextuality of PM scenarios. Typically, these rely not only on the statistics of an operational theory but also on a pre-defined *generalised probability theory* (GPT) model describing it. With these at hand, one can use linear programming to decide whether a noncontextual ontological model exists [18–20]. Conversely, in Refs. [21, 22], a bottom-up, statistics-first framework was developed to tackle the same problem. In this framework, the empirical data of a theory are collected in a matrix C of *conditional-outcome-probability evaluations* (COPE matrix). It was shown in [21, 22] that contextuality of a PM scenario relies on a single rank-based criterion of C, without reference to an underlying GPT.

A recent study by Schmid *et al.* [23] characterized the contextuality of generalized probabilistic theories (GPTs) in the prepare-transform-measure scenario (PTM) with a single transformation stage. Similar to typical PM scenario methods, this relied on an underlying GPT model and linear programming, and the composition of multistage transformations was also not addressed. In this work, our goal is to analyze generalized noncontextuality in scenarios that go beyond the simple prepare-measure case. In particular, we aim to extend the linear-algebraic framework of Refs. [21, 22] to show that a statistics-first approach is also viable in scenarios with transformations. This approach differs from Schmid *et al.* in two key respects: we are able to determine the contextuality of scenarios with an arbitrary number of transformations, and we only require the statistics of the operational theory to do so.

We begin by showing that the rank-based characterization of PM scenarios [21, 22] can be extended to PTM ones, if the COPE matrix C is replaced by an analogous COPE tensor.

We then further extend the study of contextuality to

scenarios with higher-order transformations, which have never been studied in the framework of generalized contextuality. Specifically, these are operations which act on sets of lower level transformations. Despite the more elaborate structure introduced by such operations, we show that our rank criterion is still capable of detecting contextuality in these scenarios. From a computational perspective, however, it is not immediately obvious that the criterion can be checked efficiently in this more general setting. We therefore also perform a complexity analysis to confirm that the complexity is similar to the PM case.

Next, we show that higher-order transformations can be used to obtain a contextuality condition for sequential scenarios with first-order transformations. This is achieved by defining a mapping from sequential transformation scenarios to an associated higher-order description, which captures the relevant compositional constraints. We can then check the rank criterion of the higher-order scenario. The condition is necessary-but-not-sufficient for the existence of a generalized-noncontextual ontological model for sequential PTM scenarios. As a one-sided test, it avoids the full decision problem for the contextuality of the multistage sequential model, and can therefore be checked substantially faster.

We illustrate these methods on a range of worked examples. First, we recover the expected behaviour for Spekkens' toy model, as well as an 8-state model for single-qubit stabilisers with a single stage of transformations. For the latter, the rank criterion detects the presence of transformation contextuality only, corroborating a result in [24]. We also contribute a 4-state model of a 2D toy theory that exhibits contextuality only when multiple stages of transformations are present, demonstrating that sequential structure plays a nontrivial role in witnessing contextuality.

Overall, our results provide a general framework for detecting contextuality in operational scenarios that go beyond the prepare-measure setting, including multistage and higher-order transformation structures. We view this as a step toward a more circuit-level understanding of nonclassicality, and connecting operational notions of contextuality with known requirements for quantum advantage in information processing. From a foundational perspective, we believe the work also helps clarify which features of contextuality are genuinely compositional.

The paper is organised as follows. In Sec. II we give an overview of the preliminary concepts and definitions, including operational theories and the COPE tensor formalism. In III, we introduce the different types of underlying models of operational theories, including noncontextual ontological models. In Sec. III B 3, we explain our method of deciding if a noncontexual ontological model exists for a single-stage $\mathsf{PT}^1\mathsf{M}$ scenario using our formalism, and constructing one if it does. In Sec. V, we extend this to scenarios with higher-order transformations. In Sec. IV, we show how our method from the previous section can be used to derive a noncontextuality condition in scenarios with sequential first-order transformations. Finally, in Sec. VI, we demonstrate toy models with transformations that exhibit contextuality. Discussions and conclusions are presented in Sec. VII.

II. Formalism

The groundwork for the fundamental concept in this paper, namely representation-free *operational theories*, has been laid out in Refs. [21, 22]. In this picture, each operational theory has three structures: an *operational* one, a *causal* one (including sequential and parallel compositions), and a *probabilistic* one. In this section, we recap these structures, focusing primarily on the causal structure and, above all, on the sequential composition of transformations.

A. Operational structure

The primitive elements of operational theories are lists of laboratory instructions. For example, "turn the laser on AND adjust its intensity as such" is an instruction for preparing a photonic system. Similarly, propositions such as "turn the detector on AND align its aperture as such" associated with measurements of the system are measurement instructions. Intermediate propositions such as "add a polarizer at this point in the photon path" are instructions for transforming such a system, and "rotate the polarizer's angle as such" is an instruction for applying a higher-order transformation to such a transformation.

The scenario we consider is the PTM, denoted here by $\mathsf{PT}^1\mathsf{M}$, consisting of a preparation, a single transformation, and a measurement. The operational structure thus consists of the prescriptions (propositions) for preparing, transforming, and measuring the system, denoted by the sets $\mathcal{P} := \{\mathsf{P}_i\}$, $\mathcal{T}^1 := \{\mathsf{T}^1_i\}$, and $\mathcal{M} := \{\mathsf{M}_j\}$, respectively. Later, we will extend this to an arbitrary number of transformation stages where the sets of transformation instructions are given by $\mathcal{T}^1 := \{\mathsf{T}^1_i\}$, $\mathcal{T}^2 := \{\mathsf{T}^2_i\}$, ..., $\mathcal{T}^m := \{\mathsf{T}^m_i\}$. In all scenarios, a measurement M_j results in an outcome event E^j_k from a list of possible outcomes indexed by j, $\mathcal{E}^j$.

B. Probabilistic Structure

The probabilistic structure of an operational theory of a $\mathsf{PT}^1\mathsf{M}$ scenario is provided by a probability measure p that takes every possible sentence (PTM circuit) to a real number in the unit interval. The image of p is thus the collection of conditional outcome probabilities associated with each sentence in the theory. We can organise these probabilities into a tensor C, the entries of which are the probabilities $p(k|i,j,l) := p(k|\mathsf{P}_i, \mathsf{T}^1_j, \mathsf{M}_l)$ of each outcome E^l_k, given that the system was prepared, transformed and measured according to $\mathsf{P}_i, \mathsf{T}^1_j$ and M_l re-

spectively. In a $\mathsf{PT}^1\mathsf{M}$ scenario, C is therefore a 3-tensor of *conditional outcome probabilities of events* (COPE), where we assume I preparations, J transformations, K measurements, and n_k events in the kth measurement. A COPE 3-tensor is thus the natural extension of the COPE *matrix* of a prepare-measure scenario introduced in Refs. [21, 22]. We have depicted the schematic of a COPE 3-tensor in Fig. 1. The tensor is divided into column stochastic blocks corresponding to each measurement setting, which are stacked on top of each other. Therefore, even though each probability depends on four indices, we have fused the measurement and outcome event indices into one index, which allows us to represent the COPE tensor as a 3-tensor. Each choice of transformation T^1_j thus corresponds to an $I \times \sum_{k=1}^{K} n_k$ COPE matrix,

$$C_{:j:} := \left(\begin{array}{ccc} p(1|1,j,1) & \cdots & p(1|I,j,1) \\ \vdots & & \vdots \\ p(n_1|1,j,1) & \cdots & p(n_1|I,j,1) \\ \hline & \vdots & \\ \hline p(1|1,j,K) & \cdots & p(1|I,j,K) \\ \vdots & & \vdots \\ p(n_K|1,j,K) & \cdots & p(n_K|I,j,K) \end{array}\right). \tag{1}$$

In tensor terms, each choice of dimension is called a mode. Hence, $C_{:j:}$ in the above equation is called a mode-T^1 slice. We can see that each of these slices can be thought of as representing a prepare-measure experiment, where the fixed transformation has been absorbed into either the preparations or measurements. Similarly, each mode-P slice represents a fixed choice of preparation P, and each mode-E slice represents a fixed choice of outcome event E.

The assumptions used to define the probabilistic structure of the $\mathsf{PT}^1\mathsf{M}$ scenario are the obvious extensions of those used to define the probabilistic structure of the PM scenario in Ref. [22]. In particular, we have the following.

Assumption 1. The COPE matrix contains the complete probabilistic structure of the operational theory.

In the following, we employ a 2D toy theory as a running example to clarify the concepts and approach.

Example 1. 2D Toy Theory: Probabilistic structure. Imagine a world which accommodates an *elementary* system: There are four propositions describing how to prepare the system, four propositions describing how to transform the system after preparation, and one proposition for measuring it, which yields four outcomes. We thus have $\mathcal{P} = \{\mathsf{P}_1, \mathsf{P}_2, \mathsf{P}_3, \mathsf{P}_4\}$, $\mathcal{T}^1 = \{\mathsf{T}^1_1, \mathsf{T}^1_2, \mathsf{T}^1_3, \mathsf{T}^1_4\}$ and $\mathcal{E} = \{\mathsf{E}_1, \mathsf{E}_2, \mathsf{E}_3, \mathsf{E}_4\}$. We consider the COPE tensor with mode-T^1 slices

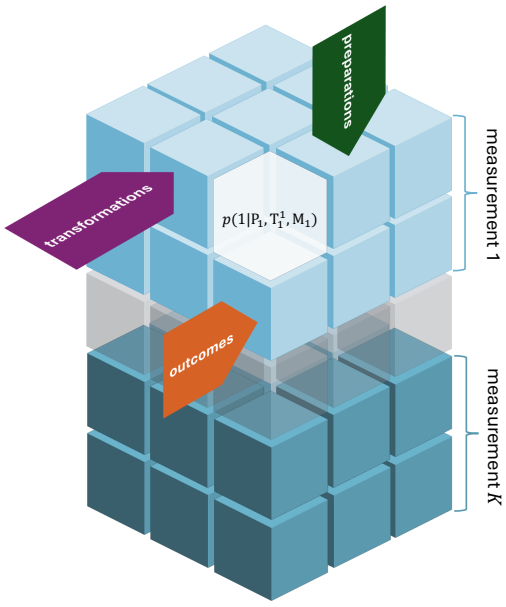

FIG. 1. Diagram of a COPE 3-tensor.

given by,

$$C^{2D}_{:1:} = \frac{1}{8}\begin{pmatrix} 1 & 3 & 3 & 1 \\ 1 & 1 & 3 & 3 \\ 3 & 1 & 1 & 3 \\ 3 & 3 & 1 & 1 \end{pmatrix}$$

$$C^{2D}_{:2:} = \frac{1}{8}\begin{pmatrix} 3 & 3 & 1 & 1 \\ 1 & 3 & 3 & 1 \\ 1 & 1 & 3 & 3 \\ 3 & 1 & 1 & 3 \end{pmatrix}$$

$$C^{2D}_{:3:} = \frac{1}{8}\begin{pmatrix} 3 & 1 & 1 & 3 \\ 3 & 3 & 1 & 1 \\ 1 & 3 & 3 & 1 \\ 1 & 1 & 3 & 3 \end{pmatrix} \tag{2}$$

$$C^{2D}_{:4:} = \frac{1}{8}\begin{pmatrix} 1 & 1 & 3 & 3 \\ 3 & 1 & 1 & 3 \\ 3 & 3 & 1 & 1 \\ 1 & 3 & 3 & 1 \end{pmatrix}$$

We note that each slice is given by cyclic shifts of the rows of the previous slice. Note that all rows and columns of each slice are convexly independent as required by a COPE matrix [22].

From the probabilistic structure, we can define a notion of operational equivalence of procedures. The operational equivalence ($\simeq$) of preparations (measurement events) means that there are no combinations of transformations and measurement events (preparations) in the

operational theory that can distinguish them [1], i.e.,

$$\begin{aligned}&\mathsf{P}_i \simeq \mathsf{P}_j \text{ iff } p(\mathsf{E}_k|\mathsf{P}_i,\mathsf{T}^1,\mathsf{M}) = p(\mathsf{E}_k|\mathsf{P}_j,\mathsf{T}^1,\mathsf{M}) \quad \forall \mathsf{E}_k,\mathsf{M},\mathsf{T}^1,\\ &\mathsf{E}_k \simeq \mathsf{E}_l \text{ iff } p(\mathsf{E}_k|\mathsf{P},\mathsf{T}^1,\mathsf{M}) = p(\mathsf{E}_l|\mathsf{P},\mathsf{T}^1,\mathsf{M}) \quad \forall \mathsf{P},\mathsf{T}^1.\end{aligned} \tag{3}$$

Similarly, two transformations are operationally equivalent iff there are no combinations of preparations and measurements that can distinguish them, i.e.,

$$\mathsf{T}^1_i \simeq \mathsf{T}^1_j \text{ iff } p(\mathsf{E}_q|\mathsf{P},\mathsf{T}^1_i,\mathsf{M}) = p(\mathsf{E}_q|\mathsf{P},\mathsf{T}^1_j,\mathsf{M}) \quad \forall \mathsf{P},\mathsf{E}_q,\mathsf{M}. \tag{4}$$

It follows that two operationally equivalent procedures give rise to two identical slices in the theory's COPE tensor. Then, Assumption 1 leads to the following observation.

Corollary 1. *For two identical slices of the COPE tensor of an operational theory, corresponding to two choices of the same procedure type, there exist no combinations of the other procedure types that can separate them.*

Other remarks in the definition of the COPE matrix of a PM scenario detailed in Ref. [21] extend to PT^{1}M scenarios in the obvious way.

Example 2. 2D Toy Theory: Operational equivalences. As an example of operational equivalences observable in the COPE tensor, consider the slices in Example 1. As usual, we assume that the set of available transformations is closed under convex combinations. Thus, we have,

$$\frac{1}{2}(C^{2D}_{:1:} + C^{2D}_{:3:}) = \frac{1}{2}(C^{2D}_{:2:} + C^{2D}_{:4:}), \tag{5}$$

implying that,

$$\frac{1}{2}(\mathsf{T}^1_1 + \mathsf{T}^1_3) \simeq \frac{1}{2}(\mathsf{T}^1_2 + \mathsf{T}^1_4). \tag{6}$$

III. Models of COPE Tensors

To construct a model that reproduces the probabilistic structure of an operational theory, mathematical objects must be assigned to the experimental procedures. Every such an assignment is called a model. In this section, we introduce three different model types and analyze their relationships. These models are used to study the nonclassicality of the operational theory.

A. preGPT and GPT Models of PT^{1}M Scenario

The first two models we consider are preGPTs and GPTs, which we extend from Ref. [21]. A preGPT for a PT^{1}M scenario is the tuple $(\mathcal{E}_{\text{pGPT}}, \mathcal{T}_{\text{pGPT}}, \mathcal{S}_{\text{pGPT}}, \mathcal{V}, \langle\cdot,\cdot\rangle)$ where $\mathcal{V}$ is an ordered inner-product vector space with $\langle\cdot,\cdot\rangle$ the inner product, and $\mathcal{E}_{\text{pGPT}}$, $\mathcal{T}_{\text{pGPT}}$ and $\mathcal{S}_{\text{pGPT}}$ are its sets of effects, transformations, and states, respectively, constructed as follows.

Each possible event of the jth measurement, E^j_k, is represented by a vector $\mathcal{K}_{\mathsf{E}}(\mathsf{E}^j_k)$, the collections of which for a fixed measurement form a *probability vector-valued measure* (PVVM) satisfying

$$\begin{aligned}&\forall \mathsf{E}_k \in \mathcal{E} \quad \mathcal{K}_{\mathsf{E}}(\mathsf{E}^j_k) \geqslant 0,\\ &\exists! u \in \mathcal{V} \text{ s.t. } \sum_k \mathcal{K}_{\mathsf{E}}(\mathsf{E}^j_k) = u,\\ &\forall j \quad \mathcal{K}_{\mathsf{E}}(\cup_k \mathsf{E}^j_k) = \sum_k \mathcal{K}_{\mathsf{E}}(\mathsf{E}^j_k).\end{aligned} \tag{7}$$

Each vector $\mathcal{K}_{\mathsf{E}}(\mathsf{E}^j_k)$ is called an effect, u is called the unit element of the preGPT, and the collection of all operationally legitimate effects is denoted by $\mathcal{E}_{\text{pGPT}}$. Further, by the convexity of the set of all events $\mathcal{E}$ and the linearity of the map $\mathcal{K}_{\mathsf{E}}$, it is assumed that $\mathcal{E}_{\text{pGPT}}$ is convex. We denote the set of extremal points of $\mathcal{E}_{\text{pGPT}}$ by ex$\mathcal{E}_{\text{pGPT}}$.

Similarly, let $\mathcal{K}_{\mathsf{P}}$ be the function that maps preparations of the operational theory to linear functionals in $\mathcal{V}^*$, the dual of $\mathcal{V}$. These vectors are called the preGPT states. It is assumed that the states are normalized in the sense that $\langle u, \mathcal{K}_{\mathsf{P}}(\mathsf{P})\rangle = 1$. Finally, let $\mathcal{K}_{\mathsf{T}^1}$ be the function that maps transformations of the operational theory to linear transformations over $\mathcal{V}^*$. We assume that the state space is invariant under $\mathcal{K}_{\mathsf{T}^1}$, that is, $\mathcal{K}_{\mathsf{T}^1}(\mathsf{T}^1_j)[\mathcal{K}_{\mathsf{P}}(\mathsf{P}_i)]$ is a valid state for every transformation T^1_j and every preparation P_i.

We denote the GPT states, effects and transformations by,

$$\begin{aligned}\mathcal{K}_{\mathsf{P}}(\mathsf{P}_i) &= s^i,\\ \mathcal{K}_{\mathsf{E}}(\mathsf{E}^l_k) &= e^{lk},\\ \mathcal{K}_{\mathsf{T}^1}(\mathsf{T}^1_j) &= \tilde{T}^{1j},\end{aligned} \tag{8}$$

respectively. To clarify, we use upper indices to label elements of the operational theory, with the lower acting as matrix indices. For instance, T^{1j}_{ik} is the element at position (i,k) in the matrix T^{1j} representing the jth transformation in the first stage.

Then, the probability of an outcome event is given by the inner product

$$\begin{aligned}p(k|\mathsf{P}_i,\mathsf{T}^1_j,\mathsf{M}_l) &= \langle \mathcal{K}_{\mathsf{E}}(\mathsf{E}^l_k), \mathcal{K}_{\mathsf{T}^1}(\mathsf{T}^1_j)[\mathcal{K}_{\mathsf{P}}(\mathsf{P}_i)]\rangle\\ &:= \langle e^{lk}, \tilde{T}^{1j}[s^i]\rangle.\end{aligned} \tag{9}$$

Note that we could also have defined transformations in terms of their action on the effect vectors. Importantly, the probability is linear in the state vector, transformation, and effect vector. Also, the maps from procedures to preGPT elements are necessarily linear. Therefore, for a procedure O_i which is probabilistically dependent on other procedures of the same kind, for instance $\mathsf{O}_i = a\mathsf{O}_j + (1-a)\mathsf{O}_k$, it is always possible to

construct a preGPT model that preserves convexity as $\mathcal{K}_{\mathsf{O}}(\mathsf{O}_i) = a\mathcal{K}_{\mathsf{O}}(\mathsf{O}_j) + (1-a)\mathcal{K}_{\mathsf{O}}(\mathsf{O}_k)$.

By choosing an orthonormal basis of $\mathcal{V}$, and the linear transformation, and using the most general linear transformation of a state vector, we write the probability rule as,

$$p(k|\mathsf{P}_i, \mathsf{T}^1_j, \mathsf{M}_l) = e^{lk\top} T^{1j} s^i. \tag{10}$$

We are now in a position to extend the definition of a GPT model for PM scenarios given in Ref. [21, 22] to the PTM cases. Defining the matrices of states, $A_{i:} = s^{i\top}$, and effects, $B_{:i} = e^i$, the COPE tensor C can be decomposed into

$$C_{ijl} = A_{i:} T^{1j} B_{:l}. \tag{11}$$

Example 3. 2D Toy Theory: PreGPT Consider the COPE tensor C^{2D} of Example 1. A possible preGPT model is given by,

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 1 & 0 & 0 & -\frac{1}{2} \\ 0 & 0 & 1 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & -\frac{1}{2} \end{pmatrix},$$

$$B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{pmatrix},$$

$$T^{11} = \begin{pmatrix} \frac{5}{32} & \frac{13}{32} & \frac{11}{32} & \frac{3}{32} & -\frac{1}{16} \\ \frac{3}{32} & \frac{3}{32} & \frac{13}{32} & \frac{13}{32} & \frac{1}{16} \\ \frac{13}{32} & \frac{5}{32} & \frac{3}{32} & \frac{11}{32} & -\frac{1}{16} \\ \frac{11}{32} & \frac{11}{32} & \frac{5}{32} & \frac{5}{32} & \frac{1}{16} \\ -\frac{1}{16} & -\frac{1}{16} & \frac{1}{16} & \frac{1}{16} & \frac{1}{8} \end{pmatrix},$$

$$T^{12} = \begin{pmatrix} \frac{5}{12} & \frac{5}{12} & \frac{1}{12} & \frac{1}{12} & -\frac{1}{12} \\ \frac{1}{12} & \frac{1}{3} & \frac{5}{12} & \frac{1}{6} & \frac{1}{12} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & -\frac{1}{12} \\ \frac{1}{3} & \frac{1}{12} & \frac{1}{6} & \frac{5}{12} & \frac{1}{12} \\ -\frac{1}{12} & -\frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{6} \end{pmatrix},$$

$$T^{13} = \begin{pmatrix} \frac{7}{16} & \frac{3}{16} & \frac{1}{16} & \frac{5}{16} & -\frac{1}{8} \\ \frac{5}{16} & \frac{5}{16} & \frac{3}{16} & \frac{3}{16} & \frac{1}{8} \\ \frac{3}{16} & \frac{7}{16} & \frac{5}{16} & \frac{1}{16} & -\frac{1}{8} \\ \frac{1}{16} & \frac{1}{16} & \frac{7}{16} & \frac{7}{16} & \frac{1}{8} \\ -\frac{1}{8} & -\frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{4} \end{pmatrix},$$

$$T^{14} = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\ \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & -\frac{1}{4} \\ 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix}. \tag{12}$$

The unit effect of this preGPT is given by the sum of the rows of A, which is $u = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 \end{pmatrix}$.

A common formalism in the multilinear algebra and tensor decomposition literature is to factorize tensors in Tucker form. A Tucker decomposition of a 3-tensor $C \in \mathbb{R}^{I\times J\times L}$ is

$$C = \mathcal{G} \times_1 U \times_2 V \times_3 W, \tag{13}$$

where the core tensor $\mathcal{G} \in \mathbb{R}^{r_1\times r_2\times r_3}$ and the factor matrices $U \in \mathbb{R}^{I\times r_{\mathsf{E}}}$, $V \in \mathbb{R}^{J\times r_{\mathsf{T}^1}}$, and $W \in \mathbb{R}^{L\times r_{\mathsf{P}}}$. Equivalently, in indices,

$$C_{ijl} = \sum_{\alpha=1}^{r_1}\sum_{\beta=1}^{r_2}\sum_{\gamma=1}^{r_3} g_{\alpha\beta\gamma}\, U_{i\alpha}\, V_{j\beta}\, W_{l\gamma}. \tag{14}$$

The Tucker rank of C is the triple

$$\mathrm{rank}_{\mathrm{Tucker}}(C) = (r_{\mathsf{E}}, r_{\mathsf{T}^1}, r_{\mathsf{P}}), \tag{15}$$

which equals the ranks of the mode-wise unfoldings, namely

$$\begin{aligned} \mathrm{rank}\big(C_{[\mathsf{E}]}\big) &= r_{\mathsf{E}}, \\ \mathrm{rank}\big(C_{\mathsf{T}^1}\big) &= r^1_{\mathsf{T}}, \\ \mathrm{rank}\big(C_{[\mathsf{P}]}\big) &= r_{\mathsf{P}}. \end{aligned} \tag{16}$$

Then, the preGPT model is equivalent to a Tucker decomposition of the form Eq. (14), where

$$A = U, \quad (T^{1j})_{\alpha\gamma} := \sum_{\beta} V_{j\beta}\, g_{\alpha\beta\gamma}, \quad B = W^\top. \tag{17}$$

This factorization can also equivalently be written in three flattened forms:

1. $C_{[\mathsf{E}]} = AB'$, where each possible composition of a preparation and transformation is represented by a column of B'.
2. $C_{[\mathsf{P}]} = A'B$ where each possible composition of a transformation and a measurement outcome is represented by a row of A'.
3. $C_{[\mathsf{T}]} = Y^1\Theta^1$, where $\Theta^1 = A^\top \otimes B$ is the Kronecker product of A and B, and Y^1 is such that $Y^1_{j:}$ is the row-flattening of T^{1j}.

Here, $C_{[\mathsf{E}]}$ is the mode-E matricization of C, $C_{[\mathsf{P}]}$ is its mode-P matricization, and $C_{[\mathsf{T}^1]}$ is its mode-T^1 matricization. Also, we will define $t^{1j} := Y^1_{j:}$ for convenience.

A GPT model of an operational theory is a preGPT that is quotiented under the operational equivalence relations for preparations, transformations and outcomes [22]. That is, a GPT does not allow for two indistinguishable procedures to have multiple representations, i.e.,

$$\begin{aligned} \mathsf{P}_1 \cong \mathsf{P}_2 &\Leftrightarrow s^{(\mathsf{P}_1)} = s^{(\mathsf{P}_2)}, \\ \mathsf{E}_1 \cong \mathsf{E}_2 &\Leftrightarrow e^{(\mathsf{E}_1)} = e^{(\mathsf{E}_2)}, \\ \mathsf{T}_1^k \cong \mathsf{T}_2^k &\Leftrightarrow T^{(\mathsf{T}_1^k)} = T^{(\mathsf{T}_2^k)}. \end{aligned} \tag{18}$$

For a preGPT model given by the COPE factorisations described above, the following criteria must be met for each type of procedure to be quotiented.

Theorem 2. *Consider a preGPT model of the operational theory. Lump the transformations with the preparations (measurements) to get a new set of preparations $\{\mathsf{P}'_i\}$ (measurements $\{\mathsf{M}'_i\}$), which corresponds to a new set of states $\{s'^i\}$ (effects $\{e'^i\}$). The original effects $\{e^i\}$ (states $\{s^i\}$) are quotiented iff the set $\{s'^i\}$ ($\{e'^i\}$) separates them.*

Proof. Assume for contradiction that the effects are quotiented, and yet there exist two distinct effects e^p and e^q which are not separated by elements of $\{s'^i\}$. That is, using Eq. (9),

$$\langle e^p, s'^i \rangle = \langle e^q, s'^i \rangle \ \forall i. \tag{19}$$

By definition, $\{s'^i\}$ contains all the extremal states that can be constructed by composing all possible preparation and transformation pairs available in the operational theory. Therefore, by Eq. (3), e^p and e^q represent operationally equivalent measurement outcomes. This is a contradiction, since operationally equivalent outcomes should have the same representation in a model with quotiented effects.

By symmetry, a similar argument can be applied to the states. □

Theorem 3. *In a preGPT of the operational theory, the representation of the transformations in the model is quotiented iff the states and effects together separate the transformation matrices.*

Proof. Similarly to the proof Theorem 2, assume for contradiction that the transformations are quotiented in the model, and yet there exist two different transformations T^{1p} and T^{1q} which are not separated by the states and effects. Then, using Eq. (9),

$$\langle e^i, T^{1p}[s^j] \rangle = \langle e^i, T^{1q}[s^j] \rangle \ \forall i, j. \tag{20}$$

Therefore, by Eq. (4), T^{1p} and T^{1q} represent operationally equivalent transformation procedures. This is a contradiction, since operationally equivalent transformation procedures should have the same representation in a model with quotiented transformations. □

Theorems 2 and 3 yield a necessary and sufficient condition for the preGPT to be fully quotiented, i.e., to form a GPT model, as follows.

Corollary 4. *Let $r_{\mathsf{E}} = \mathrm{rank}(C_{[\mathsf{E}]})$, $r_{\mathsf{P}} = \mathrm{rank}(C_{[\mathsf{P}]})$, and $r_{\mathsf{T}^1} = \mathrm{rank}(C_{[\mathsf{T}^1]})$ be the ranks associated with a preGPT given by the simultaneous factorizations,*

$$\begin{aligned} C_{[\mathsf{E}]} &= AB' \\ C_{[\mathsf{P}]} &= A'B \\ C_{[\mathsf{T}^1]} &= Y^1 \Theta^1. \end{aligned} \tag{21}$$

Then, the preGPT is a GPT iff,

$$\mathrm{rank}(A) = r_{\mathsf{E}}, \tag{22}$$

$$\mathrm{rank}(B) = r_{\mathsf{P}}, \tag{23}$$

$$\mathrm{rank}(Y^1) = r_{\mathsf{T}^1}. \tag{24}$$

We have listed these flattenings, their factorizations, and interpretations in Appendix A. Equations. (22), (23), and (24) represent the separating power of the pairs transformations-measurements, states-transformations, and states-measurements, for all states, measurements and transformations, respectively.

It is clear from Corollary 4 how to find such a GPT model.

Algorithm 1: GPT Construction Procedure

1. Find a rank factorization of $C_{[\mathsf{E}]}$ to obtain the factors A and B'
2. Find a rank factorization of $C_{[\mathsf{P}]}$ to obtain the factors A' and B
3. Pad the rows of A or columns of B with zeros at the end such that the row-length of A matches the column-length of B.
4. Construct the matrix $\Theta^1 = A^\top \otimes B$
5. Compute Y^1 as $C_{[\mathsf{T}^1]}\Theta^{1\dagger}$, where $\dagger$ denotes the pseudoinverse
6. **foreach** $\underline{t^{1i\dagger} = Y^1_{i:}}$ **do**
7. Construct the matrix T^{1i} by placing the first r_{P} entries of t^{1i} into the first row, the next r_{P} entries into the second row, and so on

Lemma 5. *Let $r_{\mathsf{E}} = \mathrm{rank}(C_{[\mathsf{E}]})$ and $r_{\mathsf{P}} = \mathrm{rank}(C_{[\mathsf{P}]})$. The minimum GPT dimension that can model the operational theory is $\max\{r_e, r_{\mathsf{P}}\}$, which is the maximum Tucker rank component.*

Proof. Algorithm 1 proves the existence of a GPT model of this dimension. To see that this is minimal, consider the first two steps of this algorithm. In the first step, $\mathrm{rank}(A)$ cannot be less than $\mathrm{rank}(C_{[\mathsf{E}]})$, therefore, since the effects are the rows of A, they cannot span a space of dimension less than $\mathrm{rank}(C_{[\mathsf{E}]})$. Similarly, we see from the second step that, since the states are columns of B, they cannot span a space of dimension less than $\mathrm{rank}(C_{[\mathcal{P}]})$. Therefore, the GPT dimension must be at least $\max\{r_e, r_{\mathsf{P}}\}$. □

Example 4. 2D Toy Theory: GPT A possible GPT model for C^{2D} is given by,

$$A = \frac{1}{2}\begin{pmatrix} 1 & 0 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \\ 1 & 1 & 0 \end{pmatrix},$$

$$T^{11} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix},$$

$$T^{12} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix},$$

$$T^{13} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \tag{25}$$

$$T^{14} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$B = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}.$$

The unit effect of this GPT is given by the sum of the rows of A, which is $u = \begin{pmatrix} 2 & 0 & 0 \end{pmatrix}$. Since $\text{rank}(A) = \text{rank}(B) = 3$, and each mode-$\mathsf{T}^1$ slice of C^{2D} is rank-3, the GPT states and effects are fully quotiented.

To verify that the GPT transformations are fully quotiented, we note that the first row of each of the first three mode-T^1 slices, $C^{2D}_{:j:}$, is given by,

$$\begin{aligned} C_{11:} &= \begin{pmatrix} 1 & 3 & 3 & 1 \end{pmatrix}, \\ C_{12:} &= \begin{pmatrix} 3 & 3 & 1 & 1 \end{pmatrix}, \\ C_{13:} &= \begin{pmatrix} 3 & 1 & 1 & 3 \end{pmatrix}. \end{aligned} \tag{26}$$

By inspection, these are linearly independent, therefore the span of the $C_{:j:}$ matrices is at least 3-dimensional. Then, since

$$T^{11} - T^{12} + T^{13} - T^{14} = 0, \tag{27}$$

the span of transformation matrices is at most 3-dimensional, proving that the GPT transformations are fully quotiented.

B. Ontological Models of $\mathsf{PT}^1\mathsf{M}$ Scenario

1. General Ontological Models

An ontological model provides a realistic interpretation of the operational theory through its probabilistic structure. It is described by an underlying ontic variable space Λ which we assume to be finite-dimensional, and spanned by a finite set of discrete ontic points $\{\lambda_i\}$. Here, each λ_i represents a definite state of the system, in which all of its properties are predetermined.

In this model, a preparation $\mathcal{P}_i$ prepares the system in a random ontic point according to some probability distribution $\mu^i(\lambda_j)$, representing an agent's limited knowledge of the true state of the system. Being a probability distribution, $\mu^i(\lambda_j)$ satisfies $\sum_j \mu^i(\lambda_j) = 1$. Each transformation T^1_i in the ontological model is represented by a stochastic map $\Gamma^{1i} : \Lambda \to \Lambda$. In the linear-algebraic formalism, these are represented by stochastic matrices encoding transition probabilities between ontic points, $\Gamma^{1i}(\lambda_k|\lambda_j)$, with $\sum_k \Gamma^{1i}(\lambda_k|\lambda_j) = 1$.

Finally, if the system is in the state λ_i just before the measurement M_j with possible outcomes $\{\mathsf{E}^j_k\}$, is performed, the outcome k occurs according to a probability distribution $\xi^{jk}(\lambda_i)$. The function ξ^{jk} is known as a response function and satisfies $\sum_k \xi^{jk}(\lambda_i) = 1$ for all λ_i.

Lemma 6. *An ontological model is a preGPT model with only nonnegative entries, up to a positive rescaling.*

Proof. We can represent each ontic point λ_i as the *ith* cartesian unit vector of a $|\Lambda|$-dimensional vector space. Therefore, we can represent the probability distribution μ^i as a nonnegative stochastic vector. As mentioned, we can represent Γ^{1i} as a stochastic matrix, such that Γ^{1i}_{kj} is the transition probability from λ_j to λ_k. Finally, we can represent ξ^{ij} as a nonnegative vector, where $\sum_{j|\mathsf{E}_j \in \mathsf{M}_i} \xi^{ij}$, is the vector of all ones, which is the unit response function.

This has the same structure as a preGPT. The only extra constraints are that all elements are nonnegative, and the constraint on the form of unit vector, which guarantees the correct normalization.

Given any nonnegative preGPT model, $C_{:j:} = AT^{1j}B$ one can efficiently convert this to an ontological model using positive diagonal rescalings, as follows,

$$\begin{aligned} AT^{1j}B &= (AD)(D^{-1}T^{1j}D')(D'^{-1}B) && (28) \\ &= R\Gamma^{1j}E, && (29) \end{aligned}$$

where D and D' are positive diagonal matrices which are chosen such that the column sum of AD is the total number of measurements, and $D'^{-1}B$ is column-stochastic. Then, since the column sum of $C_{:j:}$ is the total number of measurements, $D^{-1}T^{1j}D'$ will be column-stochastic. Here, R is the matrix of response functions and E is the matrix of epistemic states. □

Example 5. 2D Toy Theory: Ontological Model A possible ontological model for C^{2D} is

$$\begin{aligned} R &= \mathbb{1}_{4\times 4}, \\ \Gamma^{1j}_{ik} &= \delta^{[4]}_{i-k,\, j-1}\ , \\ E &= C^{2D}_{:1:}, \end{aligned} \tag{30}$$

where $\delta^{[4]}_{a,b} = 1$ if $a \equiv b \pmod 4$, and 0 otherwise. The unit response function is the sum of rows of R, which is the vector of all ones.

2. *Noncontextual Ontological Models of* $\mathsf{PT}^1\mathsf{M}$ *Scenario*

We are now ready to demonstrate how our formalism captures the notion of generalized noncontextuality [1] in $\mathsf{PT}^1\mathsf{M}$ scenarios. By definition, an ontological model is noncontextual precisely when it assigns identical representations to all operationally equivalent procedures, i.e.,

$$\begin{aligned} \mathsf{P}_1 \cong \mathsf{P}_2 &\Leftrightarrow \mu^{(\mathsf{P}_1)} = \mu^{(\mathsf{P}_2)}, \\ \mathsf{E}_1 \cong \mathsf{E}_2 &\Leftrightarrow \xi^{(\mathsf{E}_1)} = \xi^{(\mathsf{E}_2)}, \\ \mathsf{T}^k_1 \cong \mathsf{T}^k_2 &\Leftrightarrow \Gamma^{(\mathsf{T}^k_1)} = \Gamma^{(\mathsf{T}^k_2)}. \end{aligned} \tag{31}$$

Comparing these with Eq. (18), it follows from our framework that a noncontextual ontological model (NCOM) is simply an ontological model that is also a GPT. This means that for a $\mathsf{PT}^1\mathsf{M}$ scenario, the matrices of epistemic states E, response functions R, and row-flattened stochastic maps Γ^{1i} satisfy

$$\begin{aligned} \operatorname{rank}(R) &= \operatorname{rank}(C_{[\mathsf{E}]}), \\ \operatorname{rank}(E) &= \operatorname{rank}(C_{[\mathsf{P}]}), \\ \operatorname{rank}(G^1) &= \operatorname{rank}(C_{[\mathsf{T}^1]}). \end{aligned} \tag{32}$$

The above conditions are the generalization of the equirank conditions introduced in Ref. [22] for PM scenarios. The violation of the first, second and third line of Eq. (32) imply measurement, preparation and transformation contextuality, respectively.

Example 6. 2D toy theory: Noncontextual Ontological Model. A noncontextual ontological model for C^{2D} is

$$\begin{aligned} R = E &= \frac{1}{2}\begin{pmatrix} 1&1&0&0\\ 0&1&1&0\\ 0&0&1&1\\ 1&0&0&1 \end{pmatrix}, \\ \Gamma^{11} &= \frac{1}{2}\begin{pmatrix} 1&1&0&0\\ 0&1&1&0\\ 0&0&1&1\\ 1&0&0&1 \end{pmatrix}, \\ \Gamma^{12} &= \frac{1}{2}\begin{pmatrix} 1&0&0&1\\ 1&1&0&0\\ 0&1&1&0\\ 0&0&1&1 \end{pmatrix}, \\ \Gamma^{13} &= \frac{1}{2}\begin{pmatrix} 0&0&1&1\\ 1&0&0&1\\ 1&1&0&0\\ 0&1&1&0 \end{pmatrix}, \\ \Gamma^{14} &= \frac{1}{2}\begin{pmatrix} 0&1&1&0\\ 0&0&1&1\\ 1&0&0&1\\ 1&1&0&0 \end{pmatrix}. \end{aligned} \tag{33}$$

The unit response function is the vector of all ones as required.
To see that this model is noncontextual, observe that $\operatorname{rank}(R) = \operatorname{rank}(E) = \dim \operatorname{span}(\{\Gamma^{11}, \Gamma^{12}, \Gamma^{13}, \Gamma^{14}\}) = 3$, which matches $\operatorname{rank}(A) = \operatorname{rank}(B) = \dim \operatorname{span}(\{T^{11}, T^{12}, T^{13}, T^{14}\})$ from the GPT model in Eq. (25).

3. *Complexity of computing NCOMs in* $\mathsf{PT}^1\mathsf{M}$ *Scenario*

Equation (32) identify necessary and sufficient conditions for an ontological model to be noncontextual. In this section, we tackle the problem of finding nonnegative factorizations of the COPE tensor that satisfy these conditions. To do so, it is helpful to use a geometric interpretation of these equations.

We start with a geometric interpretation of the *real* factorization of a nonnegative matrix, adapted from Lemma 5 of Ref. [25]. For a nonnegative matrix C with a real factorization $C = AB$, the columns of B are interpreted as the vertices of a polytope $\mathcal{B}$, and the rows of A are interpreted as the facet inequalities of a polytope $\mathcal{A}$. For a column b_j of B, $Ab_j = C_{:j} \geq 0$ and $\sum_i (Ab_j)_i = \sum_i Cij = l$, so $b_j \in \mathcal{A}$. Hence all vertices of $\mathcal{B}$ lie in $\mathcal{A}$, i.e., $\mathcal{B} \subseteq \mathcal{A}$. This leads to a lemma regarding the geometric interpretation of RNMF due to Gillis and Glineur [26] as follows. Note that, here, without loss of generality, we have restricted the inner dimension to $r = \operatorname{rank}(C)$.

Lemma 7. *Let C be a matrix of rank r and column*

sum l with a real factorization $C = AB$ of inner dimension r. Define the polyhedron $\mathcal{A} = \{x \in rspan(A) | Ax \geq 0, \sum_{i=1}^{r}(Ax)_i = l\}$ and the polytope $\mathcal{B}$ whose vertices are the columns of B. Then, there exists an RNMF $C = RE$ of inner dimension k with $\text{rank}(R) = r$ *iff there exists a nested polytope $\mathcal{G} \subset \text{rspan}(A)$ with k vertices such that $\mathcal{B} \subseteq \mathcal{G} \subseteq \mathcal{A}$.*

Proof. First assume that such a $\mathcal{G}$ exists, and let G be the matrix whose columns are the vertices of $\mathcal{G}$. Clearly, $\mathcal{G} \subseteq \mathcal{A}$ implies $AG \geq 0$. Since $\mathcal{B} \subseteq \mathcal{G}$, each column of B is in the convex hull of the columns of G. Therefore, there exists a nonnegative matrix E such that $GE = B$. Then, defining $R := AG$, we have $RE = AGE = AB = C$. We know $\text{rank}(R) = r$ since $\text{rank}(A) = r$, and R has k columns, therefore, $C = RE$ is an RNMF of inner dimension k.

For the reverse direction, let $C = RE$ be an RNMF of inner dimension k. Then, since $\text{rank}(R) = \text{rank}(A)$, there exists a matrix G such that $R = AG$ and $\text{cspan}(G) = \text{rspan}(A)$. Now, we have $RE = A(GE)$. Since $\text{cspan}(GE) = \text{cspan}(G) = \text{rspan}(A)$, we can identify $GE = B$. Therefore, B is in the convex hull of the columns of G. □

We notice that an RNMF is always possible when k equals the number of vertices of $\mathcal{A}$. In this case, the intermediate polytope of the above proof, $\mathcal{G}$, is chosen to be $\mathcal{A}$, which guarantees that $\mathcal{B} \subseteq \mathcal{G}$. We call this RNMF the *extremal* RNMF with the restricted nonnegative matrix factor $\bar{R}$.

Lemma 8. *For an $m \times n$ nonnegative matrix C of rank r, there exists a rank-r nonnegative matrix $\bar{R}$, which we call the extremal factor, depending only on C. Moreover, for any RNMF $C = RE$, R can be written as $R = \bar{R}M$, for some nonnegative M. This $\bar{R}$ has $m^{O(r)}$ columns, and the complexity of computing it is at most $m^{O(r)} + poly(m, n)$.*

Proof. We will give a constructive proof of the existence of such a matrix $\bar{R}$. We can assume w.l.g. that C has a column sum of some integer number l. Given a rank factorization $C = AB$, define the polytopes $\mathcal{A}$ and $\mathcal{B}$ as in Lemma 7.

We first show that $\mathcal{A}$ is bounded. Assume for contradiction that there exists a set of points $x + ty$ in $\mathcal{A}$, for some fixed nonzero vectors x, y, and a scalar $t \geq 0$. Clearly, $\sum_{i=1}^{r}(A(x + ty))_i = l$ requires that $\sum_{i=1}^{r}(Ay)_i = 0$, for $r := \text{rank}(C)$. Therefore, given that $Ay \neq 0$, Ay contains a negative element. This means that $A(x+ty)$ contains a negative element for large enough t, which is a contradiction.

Given that $\mathcal{A}$ is bounded, any nested polytope $\mathcal{G}$ is in the convex hull of the vertices of $\mathcal{A}$. Let $\bar{G}$ be the matrix whose columns are vertices of $\mathcal{A}$. Then, any G as defined in Lemma 7 can be written as $\bar{G}M$, for some nonnegative matrix M. Therefore, defining $\bar{R} := A\bar{G}$, we have that any $R = AG$ for some G, can be written as $R = \bar{R}M$, for some nonnegative matrix M.

It is known that vertex enumeration of a polytope with m facets in r dimensions has complexity $m^{O(r)}$ and the number of vertices is $m^{O(r)}$ [27]. The complexity of computing the rank factorization $C = AB$ is $poly(m, n)$, hence the complexity result. □

Using a similar argument, if we instead restrict the rank of the second factor, we can analogously construct an extremal matrix, $\bar{E}$. The computational complexity of this construction is $n^{O(r)}$.

Corollary 9. *For an $m \times n$ nonnegative matrix C of rank r, there exists a rank-r nonnegative matrix $\bar{E}$ depending only on C. Moreover, for any RNMF of $C^\top$, $C^\top = E^\top R^\top$, E can be written as $E = M'\bar{E}$, for some nonnegative M. This $\bar{E}$ has $n^{O(r)}$ columns, and the complexity of computing it is $n^{O(r)} + poly(m, n)$.*

Now, we can use Lemma 8 and its corollary to decide if an NCOM of an operational theory exists, and to compute one if so.

Lemma 10. *It can be decided whether an operational theory with n_{P} preparations, n_{E} measurement events, n_{T^1} transformations, and maximum Tucker rank component r of its COPE tensor admits a noncontextual ontological model in time $poly((n_{\mathsf{E}}n_{\mathsf{P}})^{O(r)}, n_{\mathsf{T}^1})$.*

Proof. We first observe that for an ontological model given by response function matrix R, epistemic state matrix E, and stochastic transformation matrices $\{\Gamma^{1i}\}_{i=1}^{n_{\mathsf{T}^1}}$, the slice of the COPE tensor given by $R\Gamma^{1i}E$ can always be written in terms of extremal factors as $\bar{R}(M\Gamma^{1i}M')\bar{E}$, for some nonnegative M and M'. Therefore, the M and M' matrices can be absorbed into the Γ^{1i} matrices. The transformation $\Gamma^{1i} \to M\Gamma^{1i}M'$ is a linear transformation of Γ^{1i}. This means that if $\{\Gamma^{1i}\}_{i=1}^{n_{\mathsf{T}^1}}$ is a noncontextual representation of the transformations, then so is $\{M\Gamma^{1i}M'\}_{i=1}^{n_{\mathsf{T}^1}}$. Therefore, it can be assumed without loss of generality that $R = \bar{R}$ and $E = \bar{E}$.

Then, to decide if the COPE admits a NCOM, first find the extremal factors $\bar{R}$ and $\bar{E}$, which have at most $n_{\mathsf{E}}^{O(r)}$ columns and at most $n_{\mathsf{P}}^{O(r)}$ rows, respectively. Also, the complexity of this step is $n_{\mathsf{E}}^{O(r)} + n_{\mathsf{P}}^{O(r)}$. From these factors, construct the $(n_{\mathsf{E}}n_{\mathsf{P}})^{O(r)} \times n_{\mathsf{E}}n_{\mathsf{P}}$ matrix of input-output transitions $J^1 = R^\top \otimes E$. Now, the problem of finding a NCOM can be reduced to finding the matrix of vectorized stochastic matrices G^1 satisfying $G^1 J^1 = C_{[\mathsf{T}^1]}$ and $\text{rank}(G^1) = \text{rank}(C_{[\mathsf{T}^1]})$. Define $G^{1\prime} = C_{[\mathsf{T}^1]}J^\dagger$, where $J^\dagger$ is the pseudoinverse of J. Obviously, $\text{rank}(G') = \text{rank}(C_{[\mathsf{T}]})$. It follows that finding the matrix G is equivalent to finding a shear matrix Q of the form [25],

$$Q = \mathbb{1}_{k\times k} + \sum_{i=1}^{\rho}\sum_{j=1}^{k-\rho} D_{ij}\boldsymbol{a}^i\boldsymbol{b}^{j\top}, \tag{34}$$

such that

$$G^{1\prime}Q \geq 0, \tag{35}$$

wherein $\{\boldsymbol{a}^{1\top},...,\boldsymbol{a}^{\rho\top}\}$ is an orthonormal basis for the columns of J^1, $\{\boldsymbol{b}^1,...,\boldsymbol{b}^{(k-\rho)}\}$ is an orthonormal basis for the left kernel of J^1, ρ is the rank of J^1, k is the number of rows of J^1, and D is a real matrix to be found via optimization. It is immedite given the forms of Q and $G^{1\prime}$ that

$$G^{1\prime}QJ^1 = G^{1\prime}J^1, \tag{36}$$

$$\text{rank}(G^{1\prime}Q) = \text{rank}(C_{[\mathsf{T}^1]}), \tag{37}$$

therefore, the choice

$$G^1 = G^{1\prime}Q \tag{38}$$

is a valid solution for G^1. Assuming this choice for G^1, the conditions for this matrix D can be written as linear inequalities, therefore, it can be found via a linear program. In particular, each entry of G^1 corresponds to a linear inequality (nonnegativity) and is a linear function of $G^{1\prime}$, D and the a^i and b^j vectors. There are kn_{T^1} many of these inequalities and $\text{poly}(k, n_{\mathsf{T}^1})$ many variables, where $k = (n_{\mathsf{E}}n_{\mathsf{P}})^{O(r)}$. Therefore, the total complexity of solving this linear program is $poly((n_{\mathsf{E}}n_{\mathsf{P}})^{O(r)}, n_{\mathsf{T}^1})$. □

In fact, we can argue for a tighter complexity bound in the above proof. For procedure O, let r_{O} be the rank of the mode-O matricization of C. Then, using Lemma 8, $\bar{E}$ has $n_{\mathsf{P}}^{O(r_{\mathsf{P}})}$ rows, and $\bar{R}$ has $n_{\mathsf{E}}^{O(r_{\mathsf{E}})}$ columns. Constructing J^1 from these extremal factors yields a matrix with $\text{poly}(n_{\mathsf{P}}, n_{\mathsf{E}})^{r_{\mathsf{E}}+r_{\mathsf{P}}}$ rows. Then, the matrix G^1 would have dimensions $n_{\mathsf{T}^1} \times \text{poly}(n_{\mathsf{P}}, n_{\mathsf{E}})^{r_{\mathsf{E}}+r_{\mathsf{P}}}$. Therefore, finding a nonnegative solution for G^1 would have complexity $\text{poly}(n_{\mathsf{T}^1})\,\text{poly}(n_{\mathsf{P}}, n_{\mathsf{E}})^{r_{\mathsf{E}}+r_{\mathsf{P}}}$.

For simplicity, in higher-order scenarios in the coming sections, we will discuss complexity simply in terms of the minimum GPT dimension r, rather than the various ranks of matricizations of the COPE tensor. Nevertheless, it is worth noting that these ranks can be used to tighten the complexity bounds via similar arguments as above. This may be useful in cases where many of these ranks are small.

IV. Contextuality of $\mathsf{PT}^1\mathsf{T}^2\mathsf{M}$ Scenarios

The natural extension of the $\mathsf{PT}^1\mathsf{M}$ scenario is the $\mathsf{PT}^1\mathsf{T}^2\mathsf{M}$ in which there are two sequential stages of transformations. In this case, the probabilistic structure of the operational theory is given by a COPE 4-tensor where its last dimension encodes the recipes for the second transformation stage.

For a $\mathsf{PT}^1\mathsf{T}^2\mathsf{M}$ scenario with COPE tensor C, a GPT model is given by an effect matrix A, a state matrix B, and transformation matrices $\{T^{1j}\}$ and $\{T^{2k}\}$. Then, C admits the factorization,

$$C_{ijkl} = A_{i:}T^{1j}T^{2k}B_{:l}. \tag{39}$$

The resulting tensor can then be flattened as follows. The successive composition of T^1 and T^2 with preparations results in the equation $C_{[\mathsf{E}]} = AB''$, while their composition with measurements gives the equation $C_{[\mathsf{E}]} = A''B$. Similarly, composing the first transformation stage, T^1, with preparations yields the equation $C_{[\mathsf{T}^1\mathsf{P}]} = \bar{A}B'$ and composing the second stage, T^2, with measurements yields the equation $C_{[\mathsf{E}\mathsf{T}^2]} = A'\bar{B}$. Furthermore, we denote by Θ^1 the Kronecker product of A and B', by Θ^2 the Kronecker product of A' and B, by Y^1 the matrix of row-flattened T^{1j} matrices, and by Y^2 the matrix of row-flattened T^{2k} matrices. These result in the equations $C_{[\mathsf{T}^1]} = Y^1\Theta^2$ and $C_{[\mathsf{T}^2]} = Y^2\Theta^1$. Then, the conditions for the decomposition to correspond to a GPT can be given as,

$$\begin{aligned}
&\text{rank}(A) = \text{rank}(AB''),\\
&\text{rank}(B) = \text{rank}(A''B),\\
&\text{rank}(Y^1) = \text{rank}(Y^1\Theta^2),\\
&\text{rank}(Y^2) = \text{rank}(Y^2\Theta^1).
\end{aligned} \tag{40}$$

This set of equations, when satisfied, guarantees that the sets of states, measurement effects, and transformations at each stage are completely separated. This ensures that each is represented uniquely in the model, as required by a GPT.

Similarly to the previous section, an ontological model of this scenario is given by a response function matrix R, and epistemic state matrix E, and nonnegative stochastic transformation matrices $\{\Gamma^{1j}\}$ and $\{\Gamma^{2k}\}$ representing $\mathcal{T}^1$ and $\mathcal{T}^2$, respectively. Then, G^1 is the matrix of row-flattened Γ^{1j} matrices, and G^2 is the matrix of row-flattened Γ^{2k} matrices.

When finding a NCOM of this scenario, the extremal response function and epistemic state matrices, $\bar{R}$ and $\bar{E}$, respectively, can be defined by inputting $C_{[\mathsf{E}]}$ into Lemma 8, and $C_{[\mathsf{P}]}$ into Corollary 9. However, there is no analogous extremal G^1 matrix. To see this, suppose there exists a $\bar{G}^1$ such that for any choice of G^1, there exists a nonnegative M^3 that solves $G^1 = \bar{G}^1M^3$. Now, the rows of G^1 are the row-flattenings of a set of nonnegative matrices $\{\Gamma^{1j}\}$ which represent a set of first-order transformations, and there exists some set of nonnegative matrices $\{\bar{\Gamma}^{1j}\}$ whose row-flattenings are the rows of $\bar{G}^1$. Then, M^3 is equivalent to a general linear map from one set of matrices $\{\bar{\Gamma}^{1j}\}$ to another, $\{\Gamma^{1j}\}$. This linear transformation cannot necessarily be written as $\bar{\Gamma}^{1j} \to \Gamma^{1j}\bar{M}$, for some nonnegative $\bar{M}$. Since only linear transformations on $\{\bar{\Gamma}^{1j}\}$ of this form can be absorbed into the Γ^{2k} matrices, there does not generally exist an extremal noncontextual ontological representation of $\mathcal{T}^1$, $\{\bar{\Gamma}^{1j}\}$, thus there exists no general $\bar{G}^1$ matrix.

Similarly, by symmetry, there also exists no general $\bar{G}^2$ matrix, therefore we cannot isolate single layer of procedures whose representation needs to be optimized, the representations of $\mathcal{T}^1$ and $\mathcal{T}^2$ must be optimized simultaneously. It follows that our linear program cannot be extended to provide a full decision procedure for the contextuality of $\mathsf{PT}^1\mathsf{T}^2\mathsf{M}$ scenarios.

A sufficient but not necessary condition for the contextuality of a $\mathsf{PT}^1\mathsf{T}^2\mathsf{M}$ scenario can, however, be derived by

considering higher-order transformations that subsume sequential transformations as shown in Sec. V B.

V. Higher-order Scenarios

A. General case

We extend our formalism to operational theories with different causal structures, rather than just one stage of transformations, in particular, the nested $\mathsf{PT}^2[\mathsf{T}^1]\mathsf{M}$ scenario. Here, $\mathsf{T}^2[\cdot]$ is a second-order transformation, meaning that it acts on the system's transformations.

In a GPT model, the representations of preparations, measurements and T^1 transformations are represented in the same way as for the $\mathsf{PT}^1\mathsf{M}$ scenario. The T^2 transformations are represented by 4-tensors, denoted T^{2i}, which map matrices to matrices. The transformed T^1 transformations, $\mathsf{T}^2[\mathsf{T}^1]$, are represented by matrices denoted T'^{1ij} given by $T'^{1ij}_{wx} = \sum_{yz} T^{2j}_{wxyz} T^{1i}_{yz}$. That is, T'^{1ij} is the *new* first-order transformation which arises after applying the jth second order to the ith first-order. We can assume without loss of generality that T^{1i} and T'^{1ij} have the same dimensions, because if not, zero padding can be used without violating the GPT conditions.

The probabilistic structure of such an operational theory is represented by a COPE 4-tensor C, given by

$$C_{ijlp} = A_{i:} T'^{1jl} B_{:p} = \sum_{wxyz} A_{iw} T^{2l}_{wxyz} T^{1j}_{yz} B_{xp}. \tag{41}$$

Like in the single transformation stage case, such a factorization is equivalent to factorizations of flattenings of C. Two such flattenings have the same form as before, given by

1. $C_{[\mathsf{E}]} = AB'$ where each possible product of a preparation and a combination of transformations is represented by a column of B'.

2. $C_{[\mathsf{P}]} = A'B$ where each possible product of a combination of transformations and a measurement outcome is represented by a row of A'.

The factorizations of the mode-T^1 and mode-T^2 flattenings, $C_{[\mathsf{T}^1]}$ and $C_{[\mathsf{T}^2]}$, respectively, are constructed below in terms of the matrix Y^1 of row-flattened T^{1i} matrices, and the matrix Y^2 of row-flattened T^{2i} tensors.

For an index q, denote by d_q the maximum of q. Then, $C_{[\mathsf{T}^1]}$ is factorized as follows:

$$\begin{aligned} u &= (y-1)\, d_z + z, \\ v &= i + d_i\big((p-1) + d_p\,(l-1)\big), \\ \Theta^2_{uv} &= \sum_w \sum_x A_{iw}\; T^{2l}_{wxyz}\; B_{xp}, \\ (Y^1)_{ju} &= T^{1j}_{yz}, \\ C_{[\mathsf{T}^1]} &= Y^1\, \Theta^2. \end{aligned} \tag{42}$$

Similarly, for $C_{[\mathsf{T}^2]}$,

$$\begin{aligned} s &= \big(((w-1)\, d_x + (x-1))\, d_y + (y-1)\big)\, d_z + z, \\ t &= i + d_i\big((j-1) + d_j\,(p-1)\big), \\ \Theta^1_{st} &= A_{iw}\; T^{1j}_{yz}\; B_{xp}, \\ (Y^2)_{ls} &= T^{2l}_{wxyz}, \\ C_{[\mathsf{T}^2]} &= Y^2\, \Theta^1. \end{aligned} \tag{43}$$

We emphasise that the above equations are included for completeness, and essentially amount to bookkeeping of indices.

We now know all the flattenings of the COPE tensor in a $\mathsf{PT}^2[\mathsf{T}^1]\mathsf{M}$ scenario for which a GPT model must provide a rank-restricted factorization. Explicitly, the conditions of a GPT model of a $\mathsf{PT}^2[\mathsf{T}^1]\mathsf{M}$ scenario are as follows:

$$\begin{aligned} \operatorname{rank}(A) &= \operatorname{rank}(C_{[E]}), \\ \operatorname{rank}(B) &= \operatorname{rank}(C_{[P]}), \\ \operatorname{rank}(Y^1) &= \operatorname{rank}(C_{[T^1]}), \\ \operatorname{rank}(Y^2) &= \operatorname{rank}(C_{[T^2]}). \end{aligned} \tag{44}$$

This allows us to construct a method to decide if the scenario admits a NCOM. Recalling that a NCOM is simply a nonnegative GPT model (up to positive rescalings), so we represent the analogous parts of the model with R, E, Γ^{1i}, Γ^{2j}, G^1, G^2, J^1 and J^2 instead of A, B, T^{1i}, T^{2j}, Y^1, Y^2, Θ^1 and Θ^2, respectively. Then, for the sake of clarity, a NCOM must satisfy,

$$\operatorname{rank}(R) = \operatorname{rank}(C_{[E]}), \tag{45}$$

$$\operatorname{rank}(E) = \operatorname{rank}(C_{[P]}), \tag{46}$$

$$\operatorname{rank}(G^1) = \operatorname{rank}(C_{[T^1]}), \tag{47}$$

$$\operatorname{rank}(G^2) = \operatorname{rank}(C_{[T^2]}). \tag{48}$$

Also, we will use,

$$J^1_{st} = R_{iw}\; \Gamma^{1j}_{yz}\; E_{xp}, \tag{49}$$

$$J^2_{uv} = \sum_w \sum_x R_{iw}\; \Gamma^{2l}_{wxyz}\; E_{xp}, \tag{50}$$

where,

$$s = \big(((w-1)\, d_x + (x-1))\, d_y + (y-1)\big)\, d_z + z, \tag{51}$$

$$t = i + d_i\big((j-1) + d_j\,(p-1)\big), \tag{52}$$

$$u = (y-1)\, d_z + z, \tag{53}$$

$$v = i + d_i\big((p-1) + d_p\,(l-1)\big), \tag{54}$$

and as before, for an index q, d_q is the maximum of q.

Now, we are ready to prove the following theorem.

Theorem 11. *It can be decided whether an operational theory with n_{P} preparations, n_{E} measurement events, n_{T^1} first-order transformations, n_{T^2} second-order transformations and maximum Tucker rank component r of its COPE tensor admits a noncontextual ontological model in time poly$((n_{\mathsf{E}} n_{\mathsf{P}} n^1_{\mathsf{T}})^{O(r)}, n_{\mathsf{T}^2})$.*

Proof. Given a NCOM of the scenario, we can assume $R = \bar{R}$ and $E = \bar{E}$ are the extremal restricted nonnegative factors.

This is because if $R = \bar{R}M^1$ and $E = \bar{E}M^2$ for $M^1, M^2 \geq 0$, we can absorb M^1 and M^2 into the representation of $\mathsf{T}^2[\mathsf{T}^1]$ without violating any nonnegativity or rank constraints. Then, using Eq. (47), we can assume $G^1 = \bar{G}^1$, the extremal factor of $\bar{G}^1$. This is because if $G^1 = \bar{G}^1 M^3$ for $M^3 \geq 0$, M^3 represents a nonnegative linear transformation of the Γ^{1i} matrices, which can be absorbed into the representation of T^2 without violating any nonnegativity or rank constraints. Therefore, we only need to optimize the representation of T^2, i.e., the G^2 matrix.

A tensor network diagram of how the COPE tensor C is composed from R, E $\{\Gamma^{1i}\}$ and $\{\Gamma^{2j}\}$, using the representation of tensor networks introduced in Ref. [28], is given in Fig. 2. It can be seen in Fig. 2 that any nonnegative linear transformation of R, E or Γ^{1i} that acts on indices that are summed over can be instead absorbed into Γ^{2j}.

To find a solution for G^2, the process is the same as in Section III B 3. That is, define $G^{2\prime} = C_{[\mathsf{T}^2]}J^{1\dagger}$, where $\cdot^\dagger$ denotes the pseudoinverse. The next step is to search for a shear matrix Q of the form,

$$Q = \mathbb{1}_{k\times k} + \sum_{i=1}^{\rho}\sum_{j=1}^{k-\rho} D_{ij}\boldsymbol{a}^i\boldsymbol{b}^{j\top}, \tag{55}$$

such that

$$G^{2\prime}Q \geq 0, \tag{56}$$

wherein now $\{\boldsymbol{a}^{1\top}, ..., \boldsymbol{a}^{\rho\top}\}$ is an orthonormal basis for the columns of J^1, $\{\boldsymbol{b}^1, ..., \boldsymbol{b}^{(k-\rho)}\}$ is an orthonormal basis for the left kernel of J^1, ρ is the rank of J^1, k is the number of rows of J^1, and D is a real matrix to be found via optimization. Then, the choice $G^2 = G^{2\prime}Q$ is a valid solution for G^2.

Now, we consider the complexity of deciding the contextuality of a $\mathsf{PT}^2[\mathsf{T}^1]\mathsf{M}$ scenario with n_P preparations, n_E measurement events, n_{T^1} T^1 transformations, and n_{T^2} T^2 transformations. Let r be the minimum GPT dimension of the operational theory. Then, using Lemma 8, $\bar{E}$ has $n_\mathsf{P}^{O(r)}$ rows, $\bar{R}$ has $n_\mathsf{E}^{O(r)}$ columns, and $\bar{Y}^1$ has $n_{\mathsf{T}^1}^{O(r)}$ columns, therefore, by constructing J^1 from these extremal factors, it would have $\text{poly}(n_\mathsf{P}, n_\mathsf{E}, n_{\mathsf{T}^1})^r$ rows. Using this choice of J^1, the matrix G^2 would have dimensions $n_{\mathsf{T}^2} \times \text{poly}(n_\mathsf{P}, n_\mathsf{E}, n_{\mathsf{T}^1})^r$, therefore, optimizing G^2 would have complexity $\text{poly}(n_{\mathsf{T}^2})\,\text{poly}(n_\mathsf{P}, n_\mathsf{E}, n_{\mathsf{T}^1})^r$. □

The result can be extended to scenarios with even higher-order transformations in the obvious way. For example, in the $\mathsf{P}(\mathsf{T}^3[\mathsf{T}^2])[\mathsf{T}^1]\mathsf{M}$ scenario, where T^3 is third-order, the complexity would become $poly((n_\mathsf{E} n_\mathsf{P} n_{\mathsf{T}^1} n_{\mathsf{T}^2})^{O(r)}, n_{\mathsf{T}^3})$.

In this case, the flattenings can be found in the same way as before in order to find the extremal factors $\bar{E}$, $\bar{R}$, $\bar{G}^1$, $\bar{G}^2$. Then, only G^3 has to be optimized.

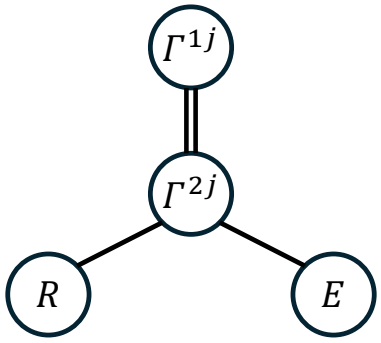

FIG. 2. Tensor-network diagram in the style of [28]: R (left) and E (right) each connect by a single leg to Γ^{2j} (center); Γ^{1i} (top) contracts with Γ^{2j} via two legs.

Of course, this can be extended to a more general scenario with k transformation stages, one stage for each order of transformation, up to kth order. This would be written as a $\mathsf{P}(...(\mathsf{T}^k[\mathsf{T}^{k-1}])...[\mathsf{T}^2])[\mathsf{T}^1]\mathsf{M}$ scenario, and the COPE of this would be a $(k{+}2)$-tensor. The complexity of deciding contextuality in this general case would be $\text{poly}(n_{\mathsf{T}^k})\,\text{poly}(n_\mathsf{P}, n_\mathsf{E}, n_{\mathsf{T}^1}, ..., n_{\mathsf{T}^{k-1}})^{(k+2)r}$, where r is the minimum GPT dimension.

B. $\mathsf{PT}^1\mathsf{T}^2\mathsf{M}$ Scenario in the Higher-Order Picture

In order to derive a contextuality condition on $\mathsf{PT}^1\mathsf{T}^2\mathsf{M}$ scenarios, we observe that they can be viewed as a special case of $\mathsf{PT}^2[\mathsf{T}^1]\mathsf{M}$ scenarios. In particular, we construct a mapping from models of $\mathsf{PT}^1\mathsf{T}^2\mathsf{M}$ scenarios to models of statistically identical higher-order scenarios.

Lemma 12. *Suppose a preGPT model of a* $\mathsf{PT}^1\mathsf{T}^2\mathsf{M}$ *scenario represents the preparation,* T^1*-transformation,* T^2*-transformation, and measurement procedures by A,* $\{T^{1i}\}$*,* $\{T^{2j}\}$*, and B, respectively. Then there exists a* $\mathsf{P}\tilde{\mathsf{T}}^2[\mathsf{T}^1]\mathsf{M}$ *scenario with the same COPE tensor, together with a preGPT model, such that A,* $\{T^{1i}\}$*, and B represent* P*,* T^1*, and* M*, respectively, and such that the representation of* $\tilde{\mathsf{T}}^2$ *is*

$$\tilde{T}^{2j}_{wxyz} = \delta_{wy}\, T^{2j}_{zx}. \tag{57}$$

Proof.

$$\begin{aligned}\sum_{yz}\tilde{T}^{2j}_{wxyz}T^{1i}_{yz} &= \sum_{yz}\delta_{wy}\,T^{2j}_{zx}T^{1i}_{yz}\\ &= \sum_{z}T^{2j}_{zx}T^{1i}_{wz}\\ &= \sum_{z}T^{1i}_{wz}T^{2j}_{zx}\\ &= (T^{1i}T^{2j})_{wx}.\end{aligned} \tag{58}$$

Therefore, under the assumed models, the two scenarios have the same COPE tensor. □

Furthermore, we use the following lemma to construct NCOMs of the higher-order scenario.

Lemma 13. *Suppose a (pre)GPT model is given for a* $\mathsf{PT}^1\mathsf{T}^2\mathsf{M}$ *scenario. Suppose also that it induces a (pre)GPT model for the associated* $\mathsf{P}\tilde{\mathsf{T}}^2[\mathsf{T}^1]\mathsf{M}$ *scenario with the same COPE, using the same representations for* P, M, *and* T^1, *and with* $\tilde{T}^{2j}_{wxyz} = \delta_{wy} T^{2j}_{zx}$. *Then the representation of* $\tilde{\mathsf{T}}^2$ *is quotiented iff the representation of* T^2 *is quotiented.*

Proof. We denote by $\mathcal{T}^2$ and $\tilde{\mathcal{T}}^2$ the sets of T^2 and $\tilde{\mathsf{T}}^2$ transformations, respectively.

Firstly, for any T^{2j}, $\mathsf{T}^{1i}\mathsf{T}^{2j}$ and $\tilde{\mathsf{T}}^{2j}[\mathsf{T}^{1i}]$ are statistically indistinguishable for all T^{1i} by assumption.

Secondly, the map $T^{2i} \mapsto \tilde{T}^{2i}$ is linear and bijective, hence it preserves linear relations. It follows that the former GPT model identifies operationally equivalent elements in span($\mathcal{T}^2$) iff the latter model identifies operationally equivalent elements in span($\tilde{\mathcal{T}}^2$). □

Now, we can derive a theorem that relates the contextuality of sequential transformations to the contextuality of higher-order transformations.

Theorem 14. *Given a* $\mathsf{PT}^1\mathsf{T}^2\mathsf{M}$ *scenario and corresponding* $\mathsf{P}\tilde{\mathsf{T}}^2[\mathsf{T}^1]\mathsf{M}$ *scenario, if the* $\mathsf{P}\tilde{\mathsf{T}}^2[\mathsf{T}^1]\mathsf{M}$ *scenario does not admit a noncontextual ontological model, then neither does the* $\mathsf{PT}^1\mathsf{T}^2\mathsf{M}$ *scenario.*

Proof. Using Lemma 13, if the $\mathsf{PT}^1\mathsf{T}^2\mathsf{M}$ scenario admits a noncontextual ontological model, then using the mapping in Lemma 12, there exists a noncontextual ontological model of the $\mathsf{P}\tilde{\mathsf{T}}^2[\mathsf{T}^1]\mathsf{M}$ scenario. Then, the result immediately follows. □

The significance of the above theorem is that it provides a nontrivial sufficient but not necessary condition for the contextuality of $\mathsf{PT}^1\mathsf{T}^2\mathsf{M}$ scenarios. The time complexity of this is at most singly exponential in the Tucker rank components and polynomial in the number of procedures in each stage. This is much faster than deciding contextuality exactly in general, which as far as we know, requires nonlinear optimization (see the conclusion of Ref. [23]). It should be noted that this condition is inequivalent to the other known sufficient but not necessary conditions for the contextuality of a $\mathsf{PT}^1\mathsf{T}^2\mathsf{M}$ scenario. One such example is shown in Ref. [23], which consists of absorbing consecutive stages of procedures into one procedure to get a new $\tilde{\mathsf{P}}\tilde{\mathsf{T}}^1\mathsf{M}$ scenario, and deciding if the resulting operational theory is contextual.

VI. Application to Known Toy Models

A. 8-state Model for Single Qubit Stabiliser

To see that the criteria of Corollary 4 certify that a model is noncontextual, we can test them on some known examples. In particular, the presence of transformation contextuality in the single qubit stabiliser subtheory has been noted in [24] and briefly in [23].

We first define the ontological model that was used in the former. Consider a single qubit on which we can perform certain stabilizer operations. The only measurements and preparations available are the Paulis $\mathcal{M} = \{X, Y, Z\}$, and their ± 1 eigenstates $\mathcal{P} = \{X^{\pm}, Y^{\pm}, Z^{\pm}\}$. In this case, the only allowed transformations are $\mathcal{T} = \{\mathsf{T}_{\mathbb{1}}, \mathsf{T}_Z, \mathsf{T}_S, \mathsf{T}_{S^{-1}}\}$, where $S = \sqrt{Z}$ is the phase gate. In particular, T_G corresponds to conjugation by the gate G, which we denote by the channel $\tau^{(G)}(\cdot) := G(\cdot)G^{\dagger}$

As usual, we call our ontic space Λ, and the size of the ontic space is $s := |\Lambda|$. The epistemic state representing the $\pm$ eigenstate of operator X will be $\mu_{X^{\pm}}(\lambda)$, and the corresponding event will be $\xi_{X^{\pm}}(\lambda)$. The channels, such as $\tau^{(G)}$, are represented by transformation matrices $\Gamma^{(H)}$.

Now, any pair of ± 1 eigenstates for a fixed Pauli are *single-shot distinguishable* (SSD), which means that there is a single measurement which discriminates them with certainty. For insance, if you know that a given state $|\psi\rangle$ belongs to the set $\{|0\rangle, \{|1\rangle\}\}$, then a measurement of Z will determine $|\psi\rangle$ with certainty. As noted in Ref. [1], any two states which are SSD must be represented by orthogonal epistemic state vectors. This is because $\sum_k \xi^k(\lambda) = 1 \ \forall \lambda$ for any complete measurement, so

$$\xi_{Z^+}(\lambda) + \xi_{Z^-}(\lambda) = 1 \ \forall \lambda.$$

Therefore, if μ_{Z^+} and μ_{Z^-} shared any support, no measurement $\{\xi_{Z^+}, \xi_{Z^-}\}$ could distinguish them with certainty.

This observation allows us to split Λ into 8 disjoint subspaces:

$$\Lambda_{x,y,z} = \Delta_{\frac{1}{2}(1+xX)} \cap \Delta_{\frac{1}{2}(1+yY)} \cap \Delta_{\frac{1}{2}(1+zZ)}, \tag{59}$$

where $(x, y, z) \in \{\pm\}^3$; the $+1$ eigenstate of X is $X^+ = \frac{1}{2}(1 + X)$ and so on. These subspaces define the shared support of any triplet of eigenstates for the X, Y, Z operators. For instance, $\mathrm{supp}(\mu_{X^+}) \cap \mathrm{supp}(\mu_{Y^+}) \cap \mathrm{supp}(\mu_{Z^+})$ is disjoint from $\mathrm{supp}(\mu_{X^-}) \cap \mathrm{supp}(\mu_{Y^+}) \cap \mathrm{supp}(\mu_{Z^-})$, etc. The reason for this is simply that, for any Pauli, the $+1$ and -1 eigenstates are SSD. For instance, a measurement of X any two states which are must have disjoint supports in a noncontextual ontological model [1]. However, we also know that a measurement of X on a Y or Z eigenstate yields $+1$ or -1 with probability $1/2$. Therefore, eigenstates of different Paulis must share some support.

Now, given that there are 8 disjoint subspaces, it is easiest to define a model with $s = 8$, since any larger model will inherit essentially the same structure. More precisely, each ontic state $\lambda = (x, y, z) \in \{\pm\}^3$ labels whether or not that ontic point is in the support of the corresponding eigenstates. For instance, choosing

$$\begin{aligned} &\lambda_1 = +++, \lambda_2 = -++, \lambda_3 = +-+, \lambda_4 = ++-, \\ &\lambda_5 = +--, \lambda_6 = -+-, \lambda_7 = --+, \lambda_8 = --- \ , \end{aligned} \tag{60}$$

then λ_1 is the unique shared support of the $+1$ eigenstates of X, Y and Z. Preparing the 1 eigenstate of X, for

instance, corresponds to choosing all ontic points where $x = +1$ [24]:

$$\mu_{X^+}(\lambda) = \begin{cases} 1/4 & \lambda \in \{\lambda_1, \lambda_3, \lambda_4, \lambda_5\} \\ 0 & \text{otherwise.} \end{cases} \tag{61}$$

Therefore, μ_{X^+} has equal support everywhere where x is $+1$, which also guarantees that in y and z it is an equal mixture of $+1$ and -1. That is, the model inherits the structure of the mutually unbiased bases it describes. The corresponding effect ξ_{X^+} is easily defined:

$$\xi_{X^+}(\lambda) = \begin{cases} 1 & \lambda \in \text{supp}(\mu_{X^+}) \\ 0 & \text{otherwise.} \end{cases} \tag{62}$$

so that $\langle \xi_{X^+}(\lambda), \mu_{X^+}(\lambda) \rangle = 1$, i.e., the probability of observing outcome + when measuring X on its + eigenstate is 1.

This assignment of epistemic states and response functions is manifestly noncontextual. The only operational equivalences are

$$\frac{1}{2}\left(X^+ + X^-\right) = \frac{1}{2}\left(Y^+ + Y^-\right) = \frac{1}{2}\left(Z^+ + Z^-\right).$$

By construction,

$$\frac{1}{2}\left(\mu_{X^+} + \mu_{X^-}\right) = \frac{1}{2}\left(\mu_{Y^+} + \mu_{Y^-}\right) = \frac{1}{2}\left(\mu_{Z^+} + \mu_{Z^-}\right)$$

clearly holds. It was noted in Ref. [24] that building preparation noncontextuality into this model inevitably leads to transformation contextuality. Our criteria corroborate this in what follows.

With descriptions of states and measurements in place, it is easy to see that conjugation by any of the gates must be represented by a permutation matrix. For instance, conjugation by Z must be represented by a permutation matrix whose action is $\Gamma^{(Z)} : (x, y, z) \to (-x, -y, z)$ and so on. No other representation is capable of replicating the statistics [24].

We first consider a $\mathsf{PT}^1\mathsf{M}$ scenario with preparations, transformations and measurements belonging to the sets $\mathcal{P}, \mathcal{T}^1$ and $\mathcal{M}$ respectively. There is only one nontrivial operational equivalence present in the set of transformations [23]:

$$\frac{1}{2}(\tau_Z(\cdot) + \tau_{\mathbb{1}}(\cdot)) \simeq \frac{1}{2}(\tau_S(\cdot) + \tau_{S^{-1}}(\cdot)), \tag{63}$$

therefore, transformation noncontextuality demands

$$\frac{1}{2}(\Gamma^{(Z)} + \Gamma^{(\mathbb{1})}) = \frac{1}{2}(\Gamma^{(S)} + \Gamma^{(S^{-1})}). \tag{64}$$

We call the 8×6 epistemic state matrix E, containing the Pauli eigenstates, and the 6×8 response function matrix R, containing the corresponding projective measurements. Using these alongside the four 8×8 transformation matrices $\Gamma^{(Z)}$, $\Gamma^{(\mathbb{1})}$, $\Gamma^{(S)}$ and $\Gamma^{(S^{-1})}$ we can form our $6 \times 4 \times 6$ COPE 3-tensor. Since we *began* with the ontological model, and used it to construct the COPE, we already have the factor matrices R and E. For the transformation factor matrix G^1, we flatten every transformation matrix into a 1×64 vector and place them into a matrix G, as described earlier. We indeed find that

$$\begin{aligned} &\text{rank}(C_{[\mathsf{E}]}) = \text{rank}(R) = 4, \\ &\text{rank}(C_{[\mathsf{P}]}) = \text{rank}(E) = 4, \\ &\text{rank}(C_{[\mathsf{T}^1]}) = 3, \\ &\text{rank}(G^1) = 4, \end{aligned} \tag{65}$$

which means that the transformations are not represented noncontextually. We emphasize as in Ref. [24], that this is a result of enforcing a noncontextual representation of preparations. Also, although we started with the ontological model, in principle the rank criteria allow us to certify contextuality *without* knowing the operational equivalence in Eq. (63).

Witnessing contextuality in this model required only $\{\mathbb{1}, Z, S, S^{-1}\}$ – indeed one could replace Z with any Pauli and S with its square-root to see the same result, as the equivalence in Eq. (63) would be unchanged. The reason that the model is contextual is because the symmetries of the Bloch sphere cannot be represented classically. Explicitly:

$$\begin{aligned} \Gamma^{(\mathbb{1})} &: (x, y, z) \to (x, y, z), \\ \Gamma^{(Z)} &: (x, y, z) \to (-x, -y, z), \\ \Gamma^{(S)} &: (x, y, z) \to (y, -x, z), \\ \Gamma^{(S^{-1})} &: (x, y, z) \to (-y, x, z), \end{aligned} \tag{66}$$

so that the equivalence in Eq. (63) is forcing a parity-preserving operation to have the same representation as a parity-reversing one.

The transformation contextuality pointed out in ref. [24] arose from a different operational equivalence, but for the same reason. Here, the authors instead considered all Paulis and the Hadamard gate, which we denote as $\mathcal{T}' = \{\mathsf{T}_{\mathbb{1}}, \mathsf{T}_X, \mathsf{T}_Y, \mathsf{T}_Z, \mathsf{T}_H\}$. Defining the depolarizing channel as:

$$\begin{aligned} \tau_D(\cdot) &= \frac{1}{4}(\mathbb{1}(\cdot)\mathbb{1} + X(\cdot)X^\dagger + Y(\cdot)Y^\dagger + Z(\cdot)Z^\dagger), \\ &= \frac{1}{4}(\tau_{\mathbb{1}}(\cdot) + \tau_X(\cdot) + \tau_Y(\cdot) + \tau_Z(\cdot)), \end{aligned} \tag{67}$$

one can see that, since it is a convex mixture of Pauli operations, it is parity-preserving. Its ontological representation must be the corresponding convex mixture:

$$\Gamma^{(D)} = \frac{1}{4}(\Gamma^{(\mathbb{1})} + \Gamma^{(X)} + \Gamma^{(Y)} + \Gamma^{(Z)}), \tag{68}$$

which is also parity-preserving. As $\tau_D(\rho) = \mathbb{1}/2 \; \forall \; \rho$, one may also define:

$$\tau_{D'}(\cdot) := H\tau_D(\cdot)H^\dagger = \tau_H(\tau_D(\cdot)),$$

$$\tau_{\mathbb{1}}(\tau_D(\cdot)) \sim \tau_H(\tau_D(\cdot)), \tag{69}$$

where clearly $\tau_{D'} = \tau_D$ since $\tau_{D'}(\rho) = \mathbb{1}/2 \ \forall \ \rho$ as well. However, the Hadamard is parity-reversing, $\Gamma^{(H)} : (x, y, z) \to (z, -y, x)$, and so $\Gamma^{(D')} = \Gamma^{(H)}\Gamma^{(D)} \neq \Gamma^{(D)}$. Once again the contextuality arises because the ontological model cannot encode equivalences between parity-reversing and preserving operations. In this case, the equivalence in Eq. (69) is defined across two stages of transformations.

With this in mind, we now define a $\mathsf{PT}^1\mathsf{T}^2\mathsf{M}$ scenario. Specifically, we form a COPE 4-tensor resulting from the same preparations and measurements, and the transformation stages $\mathcal{T}^1 = \{ \mathsf{T}_{\mathbb{1}}, \mathsf{T}_X, \mathsf{T}_Y, \mathsf{T}_Z \}$ and $\mathcal{T}^2 = \{ \mathsf{T}_{\mathbb{1}}, \mathsf{T}_H \}$. The matrices representing each transformation stage, G^1 and G^2, are formed as before by vectorising the transformation matrices in stages $\mathcal{T}^1$ and $\mathcal{T}^2$, and placing them into matrices G^1, G^2. Despite the fact that $\Gamma^{(H)}\Gamma^{(D)} \neq \Gamma^{(D)}$, we find:

$$\begin{aligned} &\text{rank}(C_{[\mathsf{E}]}) = \text{rank}(R) = 4, \\ &\text{rank}(C_{[\mathsf{P}]}) = \text{rank}(E) = 4, \\ &\text{rank}(C_{[\mathsf{T}^1]}) = \text{rank}(G_1) = 4, \\ &\text{rank}(C_{[\mathsf{T}^2]}) = \text{rank}(G_2) = 2. \end{aligned} \tag{70}$$

In other words, our rank criteria deem this scenario noncontextual. In fact, this is to be expected – used in this way, the criteria cannot recognise 'composite equivalences' of the form $\tau_H(\tau_D(\cdot)) \sim \tau_{\mathbb{1}}(\tau_D(\cdot))$. Any *distinct* choice of transformations from each stage does indeed result in a unique set of statistics, hence why G^1, G^2 and the corresponding flattenings are full rank. If, however, we place all transformations in a single stage $\mathcal{T} = \{ \mathsf{T}_{\mathbb{1}}, \mathsf{T}_X, \mathsf{T}_Y, \mathsf{T}_Z, \mathsf{T}_H, \mathsf{T}_{HX}, \mathsf{T}_{HY}, \mathsf{T}_{HZ} \}$ we find instead:

$$\begin{aligned} &\text{rank}(C_{[\mathsf{E}]}) = \text{rank}(R) = 4, \\ &\text{rank}(C_{[\mathsf{P}]}) = \text{rank}(E) = 4, \\ &\text{rank}(C_{[\mathsf{T}]}) = 6, \\ &\text{rank}(G) = 8. \end{aligned} \tag{71}$$

This implies that, among the transformations, there are two (independent) operational equivalences which are not respected at the ontological level. Alongside Eq. (69), we can also find:

$$\begin{aligned} &\frac{1}{2}(HX\rho X^\dagger H^\dagger + HZ\rho Z^\dagger H^\dagger) \equiv \frac{1}{2}(\mathbb{1}\rho\mathbb{1} + Y\rho Y^\dagger) \\ \implies &\frac{1}{2}[\tau_{HX}(\cdot) + \tau_{HZ}(\cdot)] \sim \frac{1}{2}[\tau_{\mathbb{1}}(\cdot) + \tau_Y(\cdot)] \ ; \end{aligned} \tag{72}$$

as before, the LHS is parity-reversing and the RHS is parity-preserving, so the ontological representations are not equivalent.

The fact that there is no rank separation when the stages $\mathcal{T}^1, \mathcal{T}^2$ are considered separately, but there is when they are absorbed into a single stage, implies that the contextuality arises *only* from 'composite equivalences' of the form

$$\mathsf{T}^2\mathsf{T}^1 \sim \mathsf{T}'^2\mathsf{T}'^1. \tag{73}$$

This underlines the fact that contextuality is dependent not only on the procedures available in the operational theory, but also *at what stage* each one is available; that is, which procedures are considered to be 'atomic' vs. composite. For scenarios with many transformation stages, one could use the same 'absorbing procedure' to determine whether a noncontextual model remains so when multiple stages are collapsed into one. In principle, this can also narrow down exactly where the contextuality arises. For instance if the model appears to be noncontextual except when some stages $\mathcal{T}^1, \mathcal{T}^2, \mathcal{T}^3$ are collapsed, then the equivalence which the model fails to represent must be of the form $\mathsf{T}^3\mathsf{T}^2\mathsf{T}^1 \sim \mathsf{T}'^3\mathsf{T}'^2\mathsf{T}'^1$ – here, $\mathsf{T}^i, \mathsf{T}'^i$ can be any transformations in the convex hull of those in stage $\mathcal{T}^i$.

B. Spekkens' Toy Model

Spekkens' toy model is an ontological model which mimicks many features of quantum mechanics such as entanglement, non-commutativity, interference and many more, but, notably, not contextuality [29]. It is similar to the 8-state model defined above, as it can also serve as a toy model for a single qubit. In this case, though, there are only four ontic states, $\Lambda = \{ 1, 2, 3, 4 \}$ [29, 30].

The model is *epistemically restricted* by its own kind of uncertainty principle, called the 'knowledge balance principle', which limits the amount of information an observer can gain about any system:

'If one has maximal knowledge, then for every system, at every time, the amount of knowledge one possesses about the ontic state of the system at that time must equal the amount of knowledge one lacks [29].'

For instance, suppose an observer wanted to ask questions to learn the ontic state of the system. This could be achieved with 2 questions: if the responses to *'is the state 3 or 4?'* and *'is the state 1 or 3?'* were *No* and *No*, then they would know the state was 2, for example. The knowledge balance principle prevents the observer from answering both questions. At most, they can have the answer to one of them. Therefore, there are 6 'pure' epistemic states representing maximal knowledge: they are states which have equal weight on any two ontic points. This is usually represented by disjunction; if the observer knows the state is 1 or 2 the state is written $1 \vee 2$ – in the usual notation for epistemic states we may write $\mu_{1\vee2}(\lambda) = (1/2, 1/2, 0, 0)$ and so on.

The measurements available are simply the questions we used above, of the form 'is the ontic state i or j?' For instance, 'is the ontic state 1 or 2?' can be represented by the measurement $\{ 1 \vee 2, 3 \vee 4 \}$; the first 'event' represents *Yes* and so acts like a projection onto $1 \vee 2$, the

second represents *No* and so acts like a projection onto $3 \vee 4$. Note that there are only three ways to partition Λ into two equally sized pieces, hence the observer only has 3 questions available to them.

These questions are the only way an observer can learn about the state. Since the knowledge balance principle restricts them to know the answer to only one question, the *only* 'mixed' state, where knowledge is less than maximal, is $1 \vee 2 \vee 3 \vee 4$ – this means that the observer has no information whatsoever. This can be 'decomposed' into any (disjoint) pair of pure states, for instance

$$1 \vee 2 \vee 3 \vee 4 = (1 \vee 2) \vee (3 \vee 4) = (1 \vee 3) \vee (2 \vee 4) \dots \quad (74)$$

or equivalently:

$$\begin{aligned} \mu_{1\vee2\vee3\vee4} &= (1/4, 1/4, 1/4, 1/4) \\ &= \frac{1}{2}(\mu_{1\vee2} + \mu_{3\vee4}) = \frac{1}{2}(\mu_{1\vee3} + \mu_{2\vee4}) \dots \ . \end{aligned} \quad (75)$$

In this way we can consider each disjoint pair of ontic states as a model for an orthogonal pair on the Bloch sphere, with $1 \vee 2 \vee 3 \vee 4$ as the maximally mixed state admitting multiple decompositions. In fact, these pairs of states are also a toy model for the ± 1 eigenstates of the Pauli observables, with the 'questions' corresponding the observables themselves. This is closely analogous to the 8 state model.

Another feature it shares with the 8 state model is that the allowed transformations are permutation matrices. Specifically, they are the 24 permutation matrices associated with S_4, the permutation group of 4 elements. It is useful to note that only 10 of the 4×4 transformation matrices are linearly independent. This is because every row and column must sum to 1 - this amounts to only 6 constraints (as the final row/column will be determined by the others). Considering them as vectors in $\mathbb{R}^{16}$, placed under 6 constraints, only 10 can be independent. With this in mind, we construct the COPE tensor of 6 pure states, 24 permutations and 3 yes/no questions (6 events), and find:

$$\begin{aligned} \text{rank}(C_{[\mathsf{E}]}) &= \text{rank}(R) = 4, \\ \text{rank}(C_{[\mathsf{P}]}) &= \text{rank}(E) = 4, \\ \text{rank}(C_{[\mathsf{T}]}) &= \text{rank}(G) = 10. \end{aligned} \quad (76)$$

This is to be expected; the model is manifestly noncontextual, in particular because the multiple decompositions of the mixed state are already 'built in'.

C. Information Loss and Transformation Contextuality

The operational equivalence in Eq. (69) required the depolarising channel τ_D to be defined in the ontological model. This channel also exists in Spekkens' toy model, sending every state to $1 \vee 2 \vee 3 \vee 4$; let us call it T_{D}^{toy}. Let J_n be the $n \times n$ matrix of 1s, so that $\Gamma^{(D)toy} = J_4/4$, the equal mixture of all permutation matrices, is the representation of T_{D}^{toy} in the ontological model.

It is obvious that one can manufacture a 'signal' of transformation contextuality using this channel. As an extreme example, suppose there is a stage of transformations consisting of only depolarising noise $\mathcal{T}' = \{ \mathsf{T}_{\tau_D^{toy}} \}$ amongst the set of procedures. Then the rank of every flattening of the COPE will be 1, while the ranks of the matrices comprising the ontological model will be unchanged. Naturally, this requires the observer to have constructed their ontological model *before* the depolarising stage was introduced. Therefore, one must argue that this is not an instance of contextuality. The statistics do not demand redundancy at the ontic level, but instead the redundancy is introduced by a procedure which destroys information. If the observer had *not* built an ontological model already, and was presented with completely uniform statistics (due to the depolarising stage), then building a noncontextual ontological model would be trivial.

On the other hand, this leaves open the possibility that *all* operational equivalences are the result of some loss of information. The same question was asked in Ref. [31]: 'What hypothetical physical mechanism could be responsible for the emergence of operational equivalences — the fine-tuned feature associated with contextuality?' We can ignore this on the grounds that all the instructions in an operational theory can be *implemented* by an observer. If there is a hidden process giving rise to equivalences, then by definition that process is not part of the operational theory.

D. Necessity of Sequential Scenarios

We have seen that the presence of transformation contextuality, as signaled by our rank-separation criteria, can depend on the structure of the theory. Placing different subsets of transformations in each stage can yield a noncontextual model, but placing them all in one stage can reveal contextuality. It is therefore natural to ask whether we should ever separate transformations into multiple stages, or simply place every possible transformation into a single stage.

To fully justify placing all transformations in one stage, we must ensure that there are no cases where a model is contextual over multiple stages, but noncontextual when they are collapsed into a single stage. Surprisingly, these cases *do* exist.

To see such an example, first consider the $\mathsf{PT}^1\mathsf{M}$ scenario S_1, given by Example 1, where each fixed-transformation slice of the COPE tensor is given by,

$$C^{2D}_{:1:} = \frac{1}{8} \begin{pmatrix} 1 & 3 & 3 & 1 \\ 1 & 1 & 3 & 3 \\ 3 & 1 & 1 & 3 \\ 3 & 3 & 1 & 1 \end{pmatrix} \quad (77)$$

$$C^{2D}_{:2:} = \frac{1}{8}\begin{pmatrix} 3 & 3 & 1 & 1 \\ 1 & 3 & 3 & 1 \\ 1 & 1 & 3 & 3 \\ 3 & 1 & 1 & 3 \end{pmatrix} \tag{78}$$

$$C^{2D}_{:3:} = \frac{1}{8}\begin{pmatrix} 3 & 1 & 1 & 3 \\ 3 & 3 & 1 & 1 \\ 1 & 3 & 3 & 1 \\ 1 & 1 & 3 & 3 \end{pmatrix} \tag{79}$$

$$C^{2D}_{:4:} = \frac{1}{8}\begin{pmatrix} 1 & 1 & 3 & 3 \\ 3 & 1 & 1 & 3 \\ 3 & 3 & 1 & 1 \\ 1 & 3 & 3 & 1 \end{pmatrix} \tag{80}$$

Now, we prove that the only possible noncontextual ontological model is given by:

$$R = E = \frac{1}{2}\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \tag{81}$$

$$\Gamma^{11} = \frac{1}{2}\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \tag{82}$$

$$\Gamma^{12} = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \tag{83}$$

$$\Gamma^{13} = \frac{1}{2}\begin{pmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix} \tag{84}$$

$$\Gamma^{14} = \frac{1}{2}\begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix}. \tag{85}$$

A rank factorization of $C^{2D}_{:1:}$ is given by,

$$\begin{aligned} C^{2D}_{:1:} &= \begin{pmatrix} 1 & 0 & 1 \\ 0 & -1 & 1 \\ -1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \frac{1}{8}\begin{pmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 2 & 2 & 2 & 2 \end{pmatrix} \\ &= AB \end{aligned} \tag{86}$$

Define the polytope $\mathcal{P}$ from the rows of A as $\mathcal{P} = \{x | Ax \geq 0, \sum_i x_i = 1\}$, as in Lemma. 7.We find that the extremal points are given by the columns of R. Using Lemma. 8, this means that if a noncontextual model exists, one exists with our current choice of R.

Using the same reasoning and symmetry, and Corollary. 9, a noncontextual model also exists, which uses our current choice of E. Therefore our choices of R and E are the most general. Then, with these fixed choices of R and E, it is clear from the sparsity patterns of C, R and E that noncontextuality fixes the Γ^{1i} matrices to the form given above.

Now we consider the $\mathsf{PT}^1\mathsf{T}^2\mathsf{M}$ scenario S_2 with fixed transformation slices given by,

$$C'_{:ij:} = \frac{1}{8}\begin{pmatrix} 1 & 3 & 3 & 1 \\ 1 & 1 & 3 & 3 \\ 3 & 1 & 1 & 3 \\ 3 & 3 & 1 & 1 \end{pmatrix}, \; i + j \equiv 1 \pmod 4 \tag{87}$$

$$C'_{:ij:} = \frac{1}{8}\begin{pmatrix} 3 & 3 & 1 & 1 \\ 1 & 3 & 3 & 1 \\ 1 & 1 & 3 & 3 \\ 3 & 1 & 1 & 3 \end{pmatrix}, \; i + j \equiv 2 \pmod 4 \tag{88}$$

$$C'_{:ij:} = \frac{1}{8}\begin{pmatrix} 3 & 1 & 1 & 3 \\ 3 & 3 & 1 & 1 \\ 1 & 3 & 3 & 1 \\ 1 & 1 & 3 & 3 \end{pmatrix}, \; i + j \equiv 3 \pmod 4 \tag{89}$$

$$C'_{:ij:} = \frac{1}{8}\begin{pmatrix} 1 & 1 & 3 & 3 \\ 3 & 1 & 1 & 3 \\ 3 & 3 & 1 & 1 \\ 1 & 3 & 3 & 1 \end{pmatrix}, \; i + j = 0 \pmod 4. \tag{90}$$

We see that when we absorb the two transformation stages into a single stage, the fixed-transformation slices are the same as in scenario S_1, but repeated four times each. The same is true if we absorb T^2 (T^1) into measurements (preparations): the fixed-measurement (fixed-preparation) slices are the same as in S_1, again repeated four times each. Therefore, we can use the same R and E as before, and this is still the most general choice. The Γ^{1i} matrices are therefore the only noncontextual representation of T^1. However, by symmetry, they are also the only noncontextual representation of T^2. If both T^1 and T^2 are represented with the Γ^{1i} matrices, they no longer reproduce the statistics.

Furthermore, we can show that contextuality can also be seen using Theorem 14. That is, a $\mathsf{PT}^2[\mathsf{T}^1]\mathsf{M}$ scenario with the same COPE tensor is also contextual. Firstly, as before, the most general noncontextual choice for R and E are fixed by Eq. (81) up to a permutation of rows or columns.

Then, by considering absorbing the T^2 stage into the preparations and measurements, the most general choice for T^1 is still given by the Γ^{1i} matrices in Eqs. (82)-(85).

Now, we will prove that given these representations of preparations, measurements, and T^1 transformations, we cannot have a noncontextual ontological representation of the T^2 stage. One can always follow the method in the proof of Theorem 11, however, in this case there is a shortcut.

Given that $C^{2D}_{:1:} = C'_{:ij:}$ shown in Eqs. (77) and (87) for $i + j \equiv 1 \pmod 4$, we have

$$R\Gamma^{11}E = R\Gamma'^{1ij}E, \; i + j \equiv 1 \pmod 4. \tag{91}$$

Observe that the only vectors in the right kernel of E are scalar multiples of v, where,

$$v^\top = \begin{pmatrix} 1 & -1 & 1 & -1 \end{pmatrix}. \tag{92}$$

Similarly, the only vectors in the left kernel of R are scalar multiples of $v^\top$. Therefore, Γ'^{1ij} for $i + j \equiv 1 \pmod 4$

must satisfy,

$$\Gamma'^{1ij} = \Gamma^{11} + \begin{pmatrix}\alpha_1 v & \alpha_2 v & \alpha_3 v & \alpha_4 v\end{pmatrix} + \begin{pmatrix}\beta_1 v^\top \\ \beta_2 v^\top \\ \beta_3 v^\top \\ \beta_4 v^\top\end{pmatrix}, \quad (93)$$

$$\Gamma'^{1ij} \geq 0, \quad (94)$$

where α and β are some real vectors. We show in Appendix B that this is only possible when,

$$\begin{pmatrix}\alpha_1 v & \alpha_2 v & \alpha_3 v & \alpha_4 v\end{pmatrix} + \begin{pmatrix}\beta_1 v^\top \\ \beta_2 v^\top \\ \beta_3 v^\top \\ \beta_4 v^\top\end{pmatrix} = 0, \quad (95)$$

$$\implies \Gamma'^{1ij} = \Gamma^{11},\ i + j \equiv 1 \pmod 4. \quad (96)$$

By symmetry, it is clear how this argument can be repeated to show that,

$$\Gamma'^{1ij} = \Gamma^{1l},\ i + j \equiv l \pmod 4. \quad (97)$$

In other words, the T^2 transformations simply permute the Γ^{1i} matrices. We show in Appendix C that this implies that the ontological representation of T^2 must have a 4-dimensional span. Since the rank of $C'_{[\mathsf{T}^2]}$ is only 3, this establishes rank separation and therefore contextuality. The linear program given in Section V is therefore also guaranteed to recognize contextuality in this scenario.

This example demonstrates that the contextuality of a scenario may be obscured when multiple transformation stages are absorbed into one, but can still be exposed by the linear program introduced in Section V.

VII. Conclusion

We have presented a method to decide if there exists a NCOM of an operational theory that consists of preparations, nested transformations of increasing order, and measurements. This esssentially works by iterating the linear program used to detect contextuality in single-stage transformation scenarios. The complexity of this method seems to be almost optimal. Furthermore, when such a model exists, our method computes it explicitly.

We have shown that this leads to a new sufficient-but-not-necessary condition for the existence of a NCOM model of other kinds of PTM scenarios, such as those with sequential first-order transformations. This new condition is to our knowledge much faster to verify than the fully general noncontextuality condition of a sequential PTM scenario. We then verified the condition on some known toy models, and argued that the full set of transformations allowed in an operational theory should be available at every stage, as this avoids 'false signals' of contextuality which can arise due to loss of information in noisy channels. Even with this requirement, we have given an explicit example of a PTM scenario where the sequential structure of the transformations is necessary to witness contextuality. This means that absorbing all transformation stages into one may obscure the nonclassicality of an operational theory.

A natural next step is to clarify the role of multistage (and higher-order) contextuality as a resource for information processing. Many computational models are intrinsically sequential, and it is therefore important to understand when contextuality can be detected stage-by-stage versus when it is an emergent property of composition that only appears after several transformation stages. This raises several concrete questions: can one define contextuality monotones that behave well under sequential composition and coarse-graining; and can one characterize classes of circuits for which contextuality is necessary (or sufficient) for computational speedups. Of greater practical significance is to develop robust versions of these tests that tolerate approximate operational equivalences, allowing one to certify (non)contextuality under experimentally realistic noise and finite statistics.

[1] R. W. Spekkens, Physical Review A **71** (2005).
[2] R. Kunjwal and R. W. Spekkens, Phys. Rev. Lett. **115**, 110403 (2015).
[3] E. S. Simon Kochen, Indiana Univ. Math. J. **17**, 59 (1968).
[4] S. Abramsky and A. Brandenburger, New Journal of Physics **13**, 113036 (2011).
[5] E. N. Dzhafarov and J. V. Kujala, Journal of Mathematical Psychology **74**, 11 (2016).
[6] A. Cabello, Phys. Rev. Lett. **101**, 210401 (2008).
[7] A. Cabello, S. Severini, and A. Winter, Phys. Rev. Lett. **112**, 040401 (2014).
[8] J. S. Bell, Physics Physique Fizika **1**, 195 (1964).
[9] J. S. Bell, Reviews of Modern Physics **38**, 447 (1966).
[10] C. Budroni, A. Cabello, O. Gühne, M. Kleinmann, and J.-A. Larsson, Rev. Mod. Phys. **94**, 045007 (2022).
[11] J. Bermejo-Vega, N. Delfosse, D. E. Browne, C. Okay, and R. Raussendorf, Phys. Rev. Lett. **119**, 120505 (2017).
[12] F. Shahandeh, "Quantum computational advantage implies contextuality," (2021), arXiv:2112.00024 [quant-ph].
[13] D. Schmid, H. Du, J. H. Selby, and M. F. Pusey, Phys. Rev. Lett. **129**, 120403 (2022).
[14] S. Gupta, D. Saha, Z.-P. Xu, A. Cabello, and A. S. Majumdar, Phys. Rev. Lett. **130**, 080802 (2023).
[15] D. Saha and A. Chaturvedi, Phys. Rev. A **100**, 022108 (2019).
[16] M. Howard, J. Wallman, V. Veitch, and J. Emerson, Nature **510**, 351–355 (2014).
[17] S. Mansfield and E. Kashefi, Phys. Rev. Lett. **121**,

230401 (2018).
[18] B. Craven and B. Mond, Linear Algebra and its Applications **38**, 73 (1981).
[19] J. H. Selby, E. Wolfe, D. Schmid, A. B. Sainz, and V. P. Rossi, Physical Review Letters **132** (2024).
[20] V. Gitton and M. P. Woods, Quantum **6**, 732 (2022).
[21] F. Shahandeh, T. Yianni, and M. Doosti, "Characterizing contextuality via rank separation with applications to cloning," (2024), arXiv:2406.19382 [quant-ph].
[22] F. Shahandeh, T. Yianni, and M. Doosti, "A unified linear algebraic framework for physical models and generalized contextuality," (2025), arXiv:2512.10000 [quant-ph].
[23] D. Schmid, R. D. Baldijão, J. H. Selby, A. B. Sainz, and R. W. Spekkens, "Noncontextuality inequalities for prepare-transform-measure scenarios," (2024), arXiv:2407.09624 [quant-ph].
[24] P. Lillystone, J. J. Wallman, and J. Emerson, Physical Review Letters **122**, 140405 (2019).
[25] T. Yianni and F. Shahandeh, "Complexity of contextuality," (2025), arXiv:2506.09133 [quant-ph].
[26] N. Gillis and F. Glineur, Linear Algebra and its Applications **437**, 2685 (2012).
[27] L. Khachiyan, E. Boros, K. Borys, K. Elbassioni, and V. Gurvich, Discrete & Computational Geometry **39**, 174 (2008).
[28] J. C. Bridgeman and C. T. Chubb, Journal of Physics A: Mathematical and Theoretical **50**, 223001 (2017).
[29] R. W. Spekkens, Physical Review A - Atomic, Molecular, and Optical Physics **75** (2005), 10.1103/PhysRevA.75.032110.
[30] L. Hausmann, N. Nurgalieva, and L. del Rio, (2021).
[31] L. Catani, T. D. Galley, and T. Gonda, (2024).

A. Flattenings and Factorizations

For completeness, we attach below some examples of COPE tensor flattenings and their factorizations. These apply to the PTM scenario. Here we use ξ^i, μ^i and Γ^i as shorthand for $\xi^{(\mathsf{E}_i)}, \mu^{(\mathsf{P}_i)}$ and $\Gamma^{(\mathsf{T}_i)}$. As usual the sets of effects, preparations and transformations are given by $\mathcal{E}, \mathcal{P}$ and $\mathcal{T}$.

$$C_{[\mathsf{E}]} = \begin{pmatrix} \xi^1(\lambda_1) & \xi^1(\lambda_2) & \dots & \xi^1(\lambda_s) \\ \xi^2(\lambda_1) & \xi^2(\lambda_2) & \dots & \xi^2(\lambda_s) \\ \vdots & \vdots & \dots & \vdots \\ \xi^{|\mathcal{E}|}(\lambda_1) & \xi^{|\mathcal{E}|}(\lambda_2) & \dots & \xi^{|\mathcal{E}|}(\lambda_s) \end{pmatrix} \times \tag{A1}$$

$$\sum_i \begin{pmatrix} \Gamma^1(\lambda_1|\lambda_i)\mu^1(\lambda_i) & \Gamma^2(\lambda_1|\lambda_i)\mu^1(\lambda_i) & \cdots & \Gamma^{|\mathcal{T}|}(\lambda_1|\lambda_i)\mu^1(\lambda_i) & \Gamma^1(\lambda_1|\lambda_i)\mu^2(\lambda_i) & \cdots \\ \Gamma^1(\lambda_2|\lambda_i)\mu^1(\lambda_i) & \Gamma^2(\lambda_2|\lambda_i)\mu^1(\lambda_i) & \cdots & \Gamma^{|\mathcal{T}|}(\lambda_2|\lambda_i)\mu^1(\lambda_i) & \Gamma^1(\lambda_2|\lambda_i)\mu^2(\lambda_i) & \cdots \\ \vdots & \vdots & \dots & \vdots & \vdots & \\ \Gamma^1(\lambda_s|\lambda_i)\mu^1(\lambda_i) & \Gamma^2(\lambda_s|\lambda_i)\mu^1(\lambda_i) & \cdots & \Gamma^{|\mathcal{T}|}(\lambda_s|\lambda_i)\mu^1(\lambda_i) & \Gamma^1(\lambda_s|\lambda_i)\mu^2(\lambda_i) & \cdots \\ \Gamma^1(\lambda_s|\lambda_i)\mu^1(\lambda_i) & \Gamma^2(\lambda_s|\lambda_i)\mu^1(\lambda_i) & \cdots & \Gamma^{|\mathcal{T}|}(\lambda_s|\lambda_i)\mu^1(\lambda_i) & \Gamma^1(\lambda_s|\lambda_i)\mu^2(\lambda_i) & \cdots \end{pmatrix} \tag{A2}$$

That is,

$$C_{[\mathsf{E}]} = RE' \tag{A3}$$

where E' describes the augmented set of preparations, calculated by applying every transformation in $\mathcal{T}$ to every preparation in the original set $\mathcal{P}$. The entries of $C_{[\mathsf{E}]}$ are simply the probabilities of the events described by R occurring in each of the preparations in the augmented set – essentially, a larger prepare-measure scenario. Next, we have

$$C_{[\mathsf{P}]} = \sum_i \begin{pmatrix} \xi^1(\lambda_1)\Gamma^1(\lambda_1|\lambda_i) & \xi^1(\lambda_2)\Gamma^1(\lambda_2|\lambda_i) & \cdots & \xi^1(\lambda_s)\Gamma^1(\lambda_s|\lambda_i) \\ \xi^1(\lambda_1)\Gamma^2(\lambda_1|\lambda_i) & \xi^1(\lambda_2)\Gamma^2(\lambda_2|\lambda_i) & \cdots & \xi^1(\lambda_s)\Gamma^2(\lambda_s|\lambda_i) \\ \vdots & \vdots & \dots & \vdots \\ \xi^1(\lambda_1)\Gamma^{|\mathcal{T}|}(\lambda_1|\lambda_i) & \xi^1(\lambda_2)\Gamma^{|\mathcal{T}|}(\lambda_2|\lambda_i) & \cdots & \xi^1(\lambda_s)\Gamma^{|\mathcal{T}|}(\lambda_s|\lambda_i) \\ \xi^2(\lambda_1)\Gamma^1(\lambda_1|\lambda_i) & \xi^2(\lambda_2)\Gamma^1(\lambda_2|\lambda_i) & \cdots & \xi^2(\lambda_s)\Gamma^1(\lambda_s|\lambda_i) \\ \vdots & \vdots & \dots & \vdots \end{pmatrix} \tag{A4}$$

$$\times \begin{pmatrix} \mu^1(\lambda_1) & \mu^2(\lambda_1) & \dots & \mu^{|\mathcal{P}|}(\lambda_s) \\ \mu^1(\lambda_2) & \mu^2(\lambda_2) & \dots & \mu^{|\mathcal{P}|}(\lambda_s) \\ \vdots & \vdots & \dots & \vdots \\ \mu^1(\lambda_s) & \mu^2(\lambda_s) & \dots & \mu^{|\mathcal{P}|}(\lambda_s) \end{pmatrix} \tag{A5}$$

That is,

$$C_{[\mathsf{E}]} = R'E \tag{A6}$$

where R' describes the augmented set of effects, calculated by applying every transformation *before* every effect in the original set $\mathcal{E}$. The entries of $C_{[\mathsf{P}]}$ are simply the probabilities of the events described in this augmented set occurring in any of the original preparations – again, this is just a larger prepare-measure scenario. Finally,

$$C_{[\mathsf{T}]} = \begin{pmatrix} \Gamma^1(\lambda_1|\lambda_1) & \cdots & \Gamma^1(\lambda_1|\lambda_s) & \Gamma^1(\lambda_2|\lambda_1) & \cdots \\ \Gamma^2(\lambda_1|\lambda_1) & \cdots & \Gamma^2(\lambda_1|\lambda_s) & \Gamma^2(\lambda_2|\lambda_1) & \cdots \\ \vdots & & \vdots & \vdots & \end{pmatrix} \tag{A7}$$

$$\times \begin{pmatrix} \xi^1(\lambda_1)\mu^1(\lambda_1) & \xi^1(\lambda_1)\mu^2(\lambda_1) & \cdots & \xi^1(\lambda_1)\mu^{|\mathcal{P}|}(\lambda_1) & \xi^2(\lambda_1)\mu^1(\lambda_1) & \cdots \\ \vdots & \vdots & \dots & \vdots & \vdots & \\ \xi^1(\lambda_1)\mu^1(\lambda_s) & \xi^1(\lambda_1)\mu^2(\lambda_s) & \cdots & \xi^1(\lambda_1)\mu^{|\mathcal{P}|}(\lambda_s) & \xi^2(\lambda_1)\mu^1(\lambda_s) & \cdots \\ \xi^1(\lambda_2)\mu^1(\lambda_1) & \xi^1(\lambda_2)\mu^2(\lambda_1) & \cdots & \xi^1(\lambda_2)\mu^{|\mathcal{P}|}(\lambda_1) & \xi^2(\lambda_2)\mu^1(\lambda_1) & \cdots \\ \vdots & \vdots & \dots & \vdots & \vdots & \end{pmatrix} \tag{A8}$$

That is,

$$C_{[\mathsf{T}]} = G(R^\top \otimes E) \tag{A9}$$

where G is the matrix of flattened transformation matrices. The entries of $C_{[\mathsf{T}]}$ are harder to interpret. Given a preparation and an event, they are the probabilities that the process was mediated by a given transformation.

B. Transformed Transformation Matrices in the Example

Lemma 15. *Let*

$$\begin{aligned} v &= \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}, \\ \Gamma^{11} &= \frac{1}{2}\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}. \end{aligned} \tag{B1}$$

For $\alpha = (\alpha_1, \ldots, \alpha_4)^T$ and $\beta = (\beta_1, \ldots, \beta_4)^T$, define

$$N = \Gamma^{11} + \alpha v^T + v\beta^T. \tag{B2}$$

If N is entrywise nonnegative, then

$$\alpha v^T + v\beta^T = 0. \tag{B3}$$

Equivalently, there exists $t \in \mathbb{R}$ such that

$$\alpha = t\, v \tag{B4}$$

and

$$\beta = -t\, v. \tag{B5}$$

Proof. From (B2), for all i, j,

$$N_{ij} = \Gamma^{11}_{ij} + \alpha_i v_j + v_i \beta_j. \tag{B6}$$

At each zero entry of Γ^{11}, entrywise nonnegativity implies

$$\alpha_i v_j + v_i \beta_j \geq 0. \tag{B7}$$

The zero positions of Γ^{11} are

$$(1,3), (1,4), (2,1), (2,4), (3,1), (3,2), (4,2), (4,3). \tag{B8}$$

Applying (B7) at (B8) yields

$$\begin{aligned} \alpha_1 + \beta_3 \geq 0, \qquad & -\alpha_1 + \beta_4 \geq 0, \\ \alpha_2 - \beta_1 \geq 0, \qquad & -\alpha_2 - \beta_4 \geq 0, \end{aligned} \tag{B9}$$

$$\begin{aligned} \alpha_3 + \beta_1 \geq 0, \qquad & -\alpha_3 + \beta_2 \geq 0, \\ -\alpha_4 - \beta_2 \geq 0, \qquad & \alpha_4 - \beta_3 \geq 0. \end{aligned} \tag{B10}$$

Eliminating β_j in pairs gives

$$\begin{aligned} \alpha_1 + \alpha_2 \leq 0, \qquad & \alpha_2 + \alpha_3 \geq 0, \\ \alpha_3 + \alpha_4 \leq 0, \qquad & \alpha_4 + \alpha_1 \geq 0. \end{aligned} \tag{B11}$$

Adding the first and third inequalities in (B11) and adding the second and fourth inequalities in (B11) give

$$\begin{aligned} \sum_{i=1}^{4} \alpha_i \le 0, \\ \sum_{i=1}^{4} \alpha_i \ge 0, \end{aligned} \tag{B12}$$

hence

$$\sum_{i=1}^{4} \alpha_i = 0. \tag{B13}$$

Since $(\alpha_1 + \alpha_2) \le 0$ and $(\alpha_3 + \alpha_4) \le 0$ with sum (B13), and since $(\alpha_2 + \alpha_3) \ge 0$ and $(\alpha_4 + \alpha_1) \ge 0$ with sum (B13), we obtain

$$\begin{aligned} \alpha_1 + \alpha_2 = 0, \\ \alpha_3 + \alpha_4 = 0, \\ \alpha_2 + \alpha_3 = 0, \\ \alpha_4 + \alpha_1 = 0. \end{aligned} \tag{B14}$$

Therefore

$$\begin{aligned} \alpha_2 = -\alpha_1, \\ \alpha_3 = \alpha_1, \\ \alpha_4 = -\alpha_1, \end{aligned} \tag{B15}$$

so with $t := \alpha_1$ we have

$$\alpha = t\, v. \tag{B16}$$

Substituting (B16) into (B9)–(B10) forces equalities and yields

$$\begin{aligned} \beta_1 = -t, \qquad \beta_2 = t, \\ \beta_3 = -t, \qquad \beta_4 = t, \end{aligned} \tag{B17}$$

hence

$$\beta = -t\, v. \tag{B18}$$

Finally,

$$\alpha v^T + v\beta^T = t\, vv^T - t\, vv^T = 0, \tag{B19}$$

as required. □

C. Higher-Order Transformations in the Example

Lemma 16. *Let*

$$\Gamma^{11} = \frac{1}{2} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \qquad \Gamma^{12} = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \tag{C1}$$

$$\Gamma^{13} = \frac{1}{2} \begin{pmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}, \qquad \Gamma^{14} = \frac{1}{2} \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix}, \tag{C2}$$

and let the tensors $\Gamma^{21}, \Gamma^{22}, \Gamma^{23}, \Gamma^{24} \in \mathbb{R}_{\geq 0}^{4\times 4\times 4\times 4}$ *satisfy,*

$$\sum_{y,z=1}^{4} \Gamma^{2i}_{wxyz}\, \Gamma^{1l}_{yz} = \Gamma^{1,l+i}_{wx} \quad \forall\, i, l \in \{1,2,3,4\},\ \forall\, w, x, \tag{C3}$$

where $l + i$ *is taken modulo* 4 *and relabeled into* $\{1, 2, 3, 4\}$.
Then, the span of the Γ^{2i} *tensors must be 4-dimensional.*

Proof. Firstly, observe that for any fixed matrix position (y, z) in the Γ^{1l} matrices, the position (y, z) has a nonzero entry for exactly two of the Γ^{1l} matrices, which are always consecutive. Also, observe that for any two consecutive Γ^{1l} matrices, Γ^{1i} and $\Gamma^{1,i+1}$ there exists a position (w, x) which is only nonzero on both of those matrices.
Define $r(p, q)$ such that,

$$\Gamma^{1r(p,q)}_{pq} = \Gamma^{1,r(p,q)+1}_{pq} = \frac{1}{2}. \tag{C4}$$

Now, fix (w, x, y, z) and i. By definition of $r(y, z)$ we have

$$\Gamma^{1r(y,z)}_{yz} = \Gamma^{1,r(y,z)+1}_{yz} = \tfrac{1}{2}. \tag{C5}$$

If $\Gamma^{2i}_{wxyz} > 0$, then inserting $l = r(y, z)$ into (C3) and evaluating the (w, x) entry yields a sum of nonnegative terms containing the strictly positive summand $\Gamma^{2i}_{wxyz}\Gamma^{1r(y,z)}_{yz} = \frac{1}{2}\,\Gamma^{2i}_{wxyz}$, hence

$$\Gamma^{1,r(y,z)+i}_{wx} = \tfrac{1}{2}. \tag{C6}$$

Likewise, using $l = r(y, z) + 1$ gives

$$\Gamma^{1,r(y,z)+1+i}_{wx} = \tfrac{1}{2}. \tag{C7}$$

By the definition of $r(w, x)$, this forces

$$r(w, x) \equiv r(y, z) + i \pmod 4. \tag{C8}$$

Then, taking the contrapositive and using $\Gamma^{2i}_{wxyz} \geq 0$ leads to,

$$\Gamma^{2i}_{wxyz} = 0 \quad \forall r(w, x) - r(y, z) \not\equiv i \pmod 4. \tag{C9}$$

Since the four Γ^{2i} tensors each must have nonzero entries, this implies that they have disjoint supports, therefore, they are linearly independent. Then, it clearly follows that the Γ^{2i} tensors have a 4-dimensional span. □