



HAL
open science

Formalizing contextuality in sequential scenarios

Kim Vallée, Damian Markham

► **To cite this version:**

| Kim Vallée, Damian Markham. Formalizing contextuality in sequential scenarios. 2025. <hal-05266460>

HAL Id: hal-05266460

<https://hal.science/hal-05266460v1>

Preprint submitted on 18 Sep 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-SA 4.0 - Attribution - ShareAlike - International License

Formalizing contextuality in sequential scenarios

Kim Vallée* and Damian Markham

Sorbonne University, CNRS, LIP6, F-75005 Paris, France

(Dated: September 17, 2025)

This paper provides a framework for characterizing sequential scenarios, allowing for the identification of contextuality given empirical data, and then provides precise operational interpretations in terms of the possible hidden variable model explanations of that data. Sequential scenarios are different in essence from non-local scenarios as each measurement instrument is allowed to change the state as it enters subsequent measurement instruments. Thus, it is necessary to formulate the possible state update in any hidden variable model description. Here we develop such hidden variable models for sequential scenarios, and we propose the notion of no-disturbance: an instrument A does not disturb another instrument B if the statistics of B are independent of whether A was measured or not. We define non-contextuality inequalities for the sequential scenario, and show that violation implies that the data cannot be explained by a hidden variable model that is both deterministic and not disturbing in this sense. We further provide a translation from standard contextuality frameworks to ours, providing sequential versions which carry over the same inequalities and measures of contextuality, but now with the sequential interpretations stated.

CONTENTS

I. Introduction	2	IV. Non-contextuality for sequential scenarios	9
II. Sequential scenarios	3	A. Convexity of non-contextual empirical models	10
A. Definitions	3	B. Contextual fraction for sequential scenarios	11
B. Hidden variable model	4	V. Restricted HVMs and non-contextuality	11
1. Preparation function	5	A. No-disturbance and the compatibility of marginals	12
2. Response function	5	B. No-disturbance with outcome determinism is equivalent to non-contextuality for sequential scenarios	12
3. Transfer function	5	C. No-disturbance with outcome independence is equivalent to non-contextuality for sequential scenarios	13
4. No-backward-in-time signalling	6	VI. Measurement scenario to sequential scenarios	14
5. Hidden variable model	6	A. Measurement scenario à la sheaf theoretic framework	14
C. Quantum theory as a special case	6	B. Mapping measurement scenarios to sequential scenarios	14
III. Operational considerations and HVM restrictions	7	C. Non-contextuality holds in induced sequential scenarios	15
A. Outcome determinism	7	VII. Discussion and conclusion	16
B. No-disturbance	7	A. Discussion	16
1. Example: Fair coin flip	8	B. Conclusion	17
2. Example: Non-invasive instrument	8	VIII. Acknowledgements	18
3. Example: Projective deterministic instrument	8	References	19
4. No-disturbance in quantum theory	8		
C. Outcome independence	9		
1. Example : Random resampling	9		
2. Example : Deterministic reset	9		
3. Programmable instruments and OI	9		

* kim.vallee@lip6.fr

A. Proofs	21
1. Proof of Theorem 2	21
2. Proof of Theorem 3	22
3. Proof of Theorem 4	24

I. INTRODUCTION

Quantum theory and quantum information have many applications, from communication to computation speed-up. However, developing these applications requires us to explore more fundamental features of quantum mechanics such as non-locality or contextuality. We will focus on the latter, contextuality, which takes its root in the Kochen-Specker theorem [1] but has seen renewed interest in recent years [2–6]. In order to validate or invalidate contextuality as a true property of nature, many experiments have been made [7], some of them actually using sequential measurements, i.e. measuring the system multiple times in a row without destroying it [8–11]. At the same time, formal frameworks for contextuality have been powerful tools for understanding its role and application across quantum information, from quantum computation [12, 13], to quantum metrology [14], quantum communication complexity [15] and many aspects of quantum cryptography [16, 17]. Extending formal frameworks, and formalising the interpretations, to sequential scenarios opens the door to exploring, understanding and demonstrating these non-classical effects in many more situations.

To this end, there are two issues which we address in this work. First, frameworks that relate contextuality and sequential scenarios are quite sparse [18, 19]. There have been some progress relating non-local scenarios to prepare and measure scenarios [20–22], and on prepare and measure scenario with one transformation [23, 24]. Here, we propose a general framework for sequential scenarios, and their interpretations, which builds on the ideas of the GPT and OPT formalisms [25–28] as well as the Sheaf theoretic framework for contextuality [3]. This complements the existing approaches, which either lack a clear hidden variable model, or cannot reproduce all the scenarios we allow in our framework.

The second issue is related to compatibility, or more precisely the operational assumptions required to prove contextuality. In order to experimentally check some notion of contextuality, we should work in a framework that is theory agnostic, and in the particular case of sequen-

tial instruments a common assumption is to have instrument compatibility. However, instrument compatibility has many definitions [29], and it is relatively demanding to check [27, Section 2.8.4]¹. Instead, we propose another notion called instrument no-disturbance inspired from [30], which is also operational and only requires a tomographically complete set of preparations to be checked. We note that no-disturbance, as an idea has, of course, been studied a lot, and through some formulations in certain frameworks, can be related to compatability (e.g. in OPT here [28]). Our proposed formulation is motivated by its simplicity, and as far as we know as not been formalised before in this way for general models.

We address the two previous points by giving a formal framework to study contextuality in sequential scenario with operational assumptions, and we go beyond by showing how to relate contextuality measurements scenarios [3] to the newly sequential scenarios we introduce.

We start by defining what are sequential scenarios in Section II, and what is a hidden variable model (HVM) in such sequential scenarios. We retrieve quantum theory as a special case of HVMs in our framework for sequential scenarios. We continue by introducing operational considerations and HVM restrictions in Section III, looking at outcome determinism, no-disturbance and outcome independence. Then, we introduce the formal notion of non-contextuality for sequential scenarios in IV, showing that the non-contextual set is a convex polytope and that we have a measure of contextuality, such as the contextual fraction. Here contextuality refers to the property that statistics of individual ‘contexts’ or, in our case, measurement sequences, can be understood as coming from marginals of some global probability distribution. The main point of our paper is reached in Section V, where we make a formal link between operational restrictions and contextuality. We show that contextuality, as formally defined above, is equivalent to there being an HVM that is both non-disturbing, and outcome deterministic. We also show that for scenarios where sequences are at most of length two non-contextuality can also be shown to be equivalent to there being an HVM that is non-disturbing

¹ Checking instrument compatibility would require a tomographically complete set of preparations and measurements, in addition to the instruments.

and outcome independent. Finally, we propose a mapping between measurement scenarios as described by the sheaf theoretic framework for contextuality [3] and sequential measurement scenarios in Section VI.

II. SEQUENTIAL SCENARIOS

In order to talk about features of a sequential scenario, it is necessary to give a precise description by what we consider to be sequential scenario. We clarify this notion in this section, following the steps given in the Sheaf theoretic framework for contextuality [3], more precisely by introducing the notion of a scenario followed by empirical behaviours on sequential scenarios. With these building blocks, we will have the necessary material to talk of hidden variable models and then of contextuality.

A. Definitions

In sequential scenarios, we are working with *instruments*:

Definition 1 (Instrument). *An instrument $A: \Sigma_A \rightarrow \Sigma'_A \times O_A$ takes as input a system in state $a \in \Sigma_A$ and outputs a system in a state $\sigma'_A \in \Sigma'_A$ together with a classical outcome $a \in O_A$. Where Σ_A and Σ'_A are state spaces and O_A is the set of classical outcomes allowed by the instrument.*

Instruments have had some interest in the literature, either in quantum theory [29, 30] or in GPT [25] and OPT [28] – therein called tests.

For some instrument A_2 , it might be that its outcome statistics remain unchanged whether A_1 is placed before or not. In that case we say that A_1 does not disturb A_2 :

Definition 2 (Instrument no-disturbance). *Let two instruments $A_1: \Sigma \rightarrow \Sigma \times O_{A_1}$ and $A_2: \Sigma \rightarrow \Sigma' \times O_{A_2}$, we say that A_1 does not disturb A_2 if for every outcome $a_2 \in O_{A_2}$ and every state $\sigma \in \Sigma$ we have:*

$$p_{A_2}(a_2|\sigma) = \sum_{a_1, \sigma'} p_{A_1}(a_1|\sigma) p_{A_1}(\tilde{\sigma}|a_1, \sigma) p_{A_2}(a_2|\tilde{\sigma}) \quad (\text{II.1})$$

where $p_{A_i}(a_i|\sigma)$ is the probability to have outcome a_i when A_i is performed, and the input state is σ . Similarly, $p_{A_i}(\tilde{\sigma}|a_i, \sigma)$ is the probability for

instrument A_i to prepare the state $\tilde{\sigma}$ given the input state σ and the outcome a_i .

An immediate remark about Equation (II.1) is that one could write $p_{A_1}(a_1|\sigma) p_{A_1}(\tilde{\sigma}|a_1, \sigma)$ as $p_{A_1}(\tilde{\sigma}, a_1|\sigma)$ by using the chain rule. Throughout this paper we will keep the first notation, with two terms, in order to dissociate the process of generating a new state and the process of generating an output. In order to test instrument no-disturbance, one requires *only* a tomographically complete set of preparation, since it should hold for all state. It should be contrasted with compatibility, which requires *at least* a tomographically complete set of measurements and states according to [27, Section 2.8.4].

Definition 2 deserves some additional comments regarding the existing literature. First, the term no-disturbance, or the ‘‘Gleason property’’, has sometimes been used in the literature [31, 32] to describe the compatibility of marginals, i.e.:

$$\sum_b p(a, b) = \sum_c p(a, c) = p(a).$$

It is not equivalent to our definition, and does not carry the same insights. However, it is related as we mention in Section V A. On the other hand, No Signalling In Time (NSIT) from [33] is similar to our notion of no-disturbance, but it is applied differently; NSIT should apply unconditionally to all instruments, whereas no-disturbance is assumed to hold for only some set of instruments. Finally, we do not refer here to the notion of state no-disturbance, as in the ‘‘no information without disturbance’’ principle [34, 35], but rather to outcome no disturbance – i.e. the fact that the outcome of a measurement is independent of the previous measurement.

When one has multiple sequences of instruments, this forms what we call a *sequential scenario*:

Definition 3 (Sequential scenario). *A sequential scenario is given by a tuple $\langle X, \mathcal{S}, O \rangle_{\text{seq}}$ where:*

- X is the set of instrument labels;
- \mathcal{S} the set of sequences, where a sequence $S \in \mathcal{S}$ is a set of instrument labels indexed by their ordering $S = \{A_i \mid A \in X, i \in I\}$ where $I \subset \mathbb{N}$ is an indexed set. The set of all possible sequences given a set X and an indexed set I is written \mathfrak{S}_I^X ;
- $O = (O_x)_{x \in X}$ is the set of classical outcomes for each instrument. When we consider a sequence of instruments $S \in \mathfrak{S}_I^X$,

we define O_S as the Cartesian product of the possible outcomes:

$$O_S = \prod_{\substack{s \in S, x \in X \\ i \in I \\ x_i = s}} O_x.$$

Where I is an indexed set and S an arbitrary sequence.

We point out two implicit assumptions which are made in sequential scenarios. First, all the first instruments in a sequence must have the same input state space, say Σ_{in} , and second whenever two instruments are made in a sequence they must have matching state space. As an example, let $A: \Sigma_A \rightarrow \Sigma'_A \times O_A$ and $B: \Sigma_B \rightarrow \Sigma'_B \times O_B$ such that $S = \{A_1, B_2\}$ then one must have $\Sigma'_A = \Sigma_B$. Moreover, since A_1 is the first instrument in the sequence we necessarily have $\Sigma_A = \Sigma_{\text{in}}$.

We give below an example with a sequential version of the KCBS scenario [36, 37].

Example 1 (Sequential KCBS). We define the sequential KCBS scenario $\langle X_{\text{KCBS}}, \mathcal{S}_{\text{KCBS}}, O_{\text{KCBS}} \rangle_{\text{seq}}$ where we have five instruments $X_{\text{KCBS}} = \{A^i\}_{i=0, \dots, 4}$, and they all have two possible outcomes $\forall A^i \in X: O_{\text{KCBS}, A^i} = \{0, 1\}$. Then we define the sequences as:

$$\begin{aligned} \mathcal{S}_{\text{KCBS}} = \{ & \{A_1^0, A_2^1\}, \{A_1^1, A_2^2\}, \\ & \{A_1^2, A_2^3\}, \{A_1^3, A_2^4\}, \\ & \{A_1^4, A_2^0\} \} \end{aligned} \quad (\text{II.2})$$

Where the subscript is the order inside a sequence, and the superscript defines the instrument.

Note that in Example 1 we could have decided of different order in the sequences $S \in \mathcal{S}_{\text{KCBS}}$, and that would still be a valid scenario. This illustrates that there are multiple possible sequential scenarios associated to a non-contextuality scenario. We will give more details on this possibility in Section IV.

When we run an experiment or make a model from a theory of a given sequential scenario, we obtain a set of probabilities. We call these probabilities an *empirical behaviour*, which is defined on a sequential scenario:

Definition 4 ((Sequential) Empirical behaviour). An (sequential) empirical behaviour e

is defined on a sequential scenario $\langle X, \mathcal{S}, O \rangle_{\text{seq}}$ as a family of probability distribution $e = (e_S)_{S \in \mathcal{S}}$ where each $e_S: O_S \rightarrow [0, 1]$ is a probability distribution on classical outcomes for the instruments in a sequence.

We also define the notion of compatibility of marginals, in the same spirit as [3]:

Definition 5 (Compatibility of marginals). We say that an empirical behaviour e on a sequential scenario $\langle X, \mathcal{S}, O \rangle_{\text{seq}}$ respects the compatibility of marginals if for any instrument $A \in X$, such that for all sequences $S_1, S_2 \in \mathcal{S}$ and all $k, k' \in \mathbb{N}$ where $A_k \in S_1$ and $A_{k'} \in S_2$ and $A_k, A_{k'} \in S_1 \cap S_2$ then

$$\sum_{a_i \neq a_k} p(a_1, \dots, a_{|S_1|} | S_1) = \sum_{a_j \neq a_{k'}} p(a_1, \dots, a_{|S_2|} | S_2) \quad (\text{II.3})$$

where the outcomes a_k and $a_{k'}$ correspond to the instruments A_k and $A_{k'}$ respectively.

Notably, the compatibility of marginals matches the notion of no-signalling in the case of non-local scenarios, such as Bell scenarios [3].

B. Hidden variable model

Since this paper aims at deriving an operational way to rule out some theories, it is useful to propose a model which can reproduce the statistics of any theory. Such a model is called a Hidden Variable Model (HVM), and it can for instance reproduce both classical theory but also quantum theory or any post-quantum theory. Restricting a HVM may result in a specific theory, e.g., to reproduce non-contextual theories the Kochen-Specker theorem constrains a HVM to be deterministic and parameter independent [1].

We provide in this subsection all the building blocks of a HVM for sequential scenarios, which we then combine². In the following, we will refer to an hidden variable $\lambda \in \Lambda$ where Λ is the set of all possible hidden variables. We consider Λ to be finite, but eventually arbitrarily large, to simplify notation without losing generality. In any HVM,

² In this paper we use the terminology HVM and not ontological model to match the historical notation in the sheaf theoretic framework for contextuality [3] but they should be understood as conceptually the same.

instruments are represented by $A: \Lambda_A \rightarrow \Lambda'_A \times O_A$, where $\Lambda_A \subseteq \Lambda$ and $\Lambda'_A \subseteq \Lambda$. For simplicity of notation we will always write $A: \Lambda \rightarrow \Lambda \times O_A$ for any instrument.

The general sequential scenario is represented in Figure 1, where first a hidden variable λ is prepared following a distribution $\mu_\rho(\lambda)$ given by a state preparation ρ , then measured and transformed multiple times through instruments. Each time a hidden variable λ passes through an instrument it is transformed to another hidden variable λ' , and a classical output is produced³. We formalize this procedure in this section by preparation functions, response functions and transfer functions respectively.

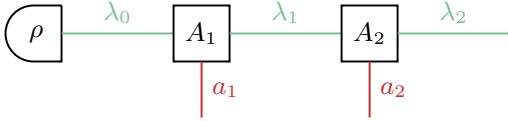


Figure 1. Example of a sequence, where the hidden variables are represented in green, and the classical outcomes in red. First there is a state preparation ρ where the hidden variable λ_0 is sampled with probability $\mu_\rho(\lambda_0)$. This hidden variable is then an input to an instrument A_1 . It has two results: one a classical outcome a_1 given with probability $\xi_{A_1}(a_1|\lambda_0)$, and the second a new hidden variable λ_1 with probability $\Gamma_{A_1}(\lambda_1|\lambda_0, a_1)$. The new hidden variable produced is forwarded to the next instrument A_2 , and it produces a classical outcome a_2 as well as a hidden variable λ_2 .

1. Preparation function

A state preparation ρ samples hidden variables with an associated probability distribution $\mu_\rho: \Lambda \rightarrow [0, 1]$ which has the properties:

$$\forall \lambda \in \Lambda: \mu_\rho(\lambda) \geq 0 \quad (\text{II.4})$$

$$\sum_{\lambda \in \Lambda} \mu_\rho(\lambda) = 1. \quad (\text{II.5})$$

The support of this probability distribution is written $\Lambda_\rho = \{\lambda | \mu_\rho(\lambda) > 0\}$. We call μ_ρ the *preparation function* of ρ .

2. Response function

Each instrument A responds to a hidden variable by outputting a classical outcome, which we denote by the lower case letter a , and we model it by a *response function* $\xi_A(\cdot|\lambda): O_A \rightarrow [0, 1]$, which is the probability to have the outcome a given the hidden variable λ when measuring A . These *response functions* are normalized – on the outcomes – and positive such that for all $\lambda \in \Lambda$:

$$\sum_a \xi_A(a|\lambda) = 1 \quad (\text{II.6})$$

$$\forall a \in O_A: \xi_A(a|\lambda) \geq 0. \quad (\text{II.7})$$

Notice that each response function is independent of the order of the instrument, this information may only be accessible through the hidden variable λ . In other words we make the following assumption:

$$\forall k \in \mathbb{N}: \xi_{A_k}(\cdot|\lambda) \equiv \xi_A(\cdot|\lambda)$$

3. Transfer function

After a given instrument A_i , at step i , the previous hidden variable λ_{i-1} is resampled in a new hidden variable λ_i , possibly stochastically. The resampling of the hidden variable is determined following the function $\Gamma_{A_i}(\cdot|\lambda_{i-1}, a_i): \Lambda \rightarrow [0, 1]$, which represents the probability to sample λ_i given the previous hidden variable λ_{i-1} and the classical outcome a_i . We naturally assume that the resampling is normalized, and it behaves as a probability distribution such that $\forall \lambda_{i-1} \in \Lambda$ and for all $a_i \in O_{A_i}$:

$$\sum_{\lambda_i \in \Lambda} \Gamma_{A_i}(\lambda_i|\lambda_{i-1}, a_i) = 1 \quad (\text{II.8})$$

$$\Gamma_{A_i}(\lambda_i|\lambda_{i-1}, a_i) \geq 0. \quad (\text{II.9})$$

As one may notice, we have decided to remove the dependency in all the past measurements from the transfer function, e.g.,

$$\Gamma_{A_i}(\lambda_i|\lambda_{i-1}, \dots, \lambda_0, a_i, \dots, a_0).$$

However, we do not lose the generality, since this can be encoded in the state λ_{i-1} as it may have arbitrarily large dimension.

³ One might note that all these notions are special case of channels [25], but we keep them separated voluntarily in order to simplify the conditions we impose on them.

4. No-backward-in-time signalling

The most general HVM cannot be built only from the preparation, transfer and response functions, only because one could allow for signalling backward in time, which is highly counter-intuitive. Consequently, in the remaining of this paper, we make the assumption of No-Backwards-in-Time-Signalling (NBTS), similar to the one described in [38]:

Definition 6 (No-Backwards-in-Time-Signalling). *Any outcome or state at time t cannot depend on any outcome or state that occurs at a later time $t + \varepsilon$, with $\varepsilon > 0$.*

We have not formally introduced the notion of time in our framework, but it is implicitly assumed from the ordering of the instruments. For instance in a sequential scenario $\langle X, \mathcal{S}, O \rangle_{\text{seq}}$, let a sequence $S \in \mathcal{S}$ of length N . Such that the first instrument is A_1 and the last is A_N . Take the outcome that occurs at time i , then by the NBTS condition:

$$\begin{aligned} p(a_i | a_1, \dots, a_N, \lambda_0, \dots, \lambda_N) \\ = p(a_i | a_1, \dots, a_{i-1}, \lambda_0, \dots, \lambda_{i-1}). \end{aligned} \quad (\text{II.10})$$

$$\forall S \in \mathcal{S}: \forall o \in O_S: \quad h_S(o) = \sum_{\lambda_0, \dots, \lambda_{|S|}} \mu(\lambda_0) \prod_{i=1}^{|S|} \xi_{A_i}(a_i | \lambda_{i-1}) \Gamma_{A_i}(\lambda_i | \lambda_{i-1}, a_i) \quad (\text{II.13})$$

In the following, we will only make use of Equation (II.13) to describe any empirical model generated by a HVM, thus ruling out any form of backward in time signalling.

⁴ This can be proven by taking the most general probability $p_S(a_1, \dots, a_{|S|}, \lambda_1, \dots, \lambda_{|S|})$, applying the chain rule, and removing any dependency in the terms that must occur at a later time.

5. Hidden variable model

Lastly, we give the definition of a hidden variable model:

Definition 7 (Hidden variable model). *A hidden variable model on a sequential scenario $\langle X, \mathcal{S}, O \rangle_{\text{seq}}$ is a tuple $\langle \Lambda, \mu, (h^\lambda)_{\lambda \in \Lambda} \rangle$ with:*

- (i) Λ is the set of hidden variables;
- (ii) $\mu: \Lambda \rightarrow [0, 1]$ is the preparation function;
- (iii) for all λ , $h^\lambda = (h_S^\lambda)_{S \in \mathcal{S}}$ is a family of probability distributions $h_S^\lambda: O_S \rightarrow [0, 1]$.

These give rise to an empirical model:

$$h := \sum_{\lambda \in \Lambda} \mu(\lambda) h^\lambda \quad (\text{II.11})$$

While definition 7 is the most general definition for a HVM, we will only be concerned with HVMs which are NBTS, which means one can rewrite any h_S^λ as⁴:

$$h_S^\lambda = \prod_{i=1}^{|S|} \xi_{A_i}(a_i | \lambda_{i-1}) \Gamma_{A_i}(\lambda_i | \lambda_{i-1}, a_i) \quad (\text{II.12})$$

Putting together Equation (II.11) with Equation (II.12) we obtain that any HVM $h = (h_S)_{S \in \mathcal{S}}$ can be written as:

C. Quantum theory as a special case

It is important to check that our HVM can at least recover quantum theory as a special case. This is possible if one assigns the set of hidden variables to the set of pure states $\Lambda = \mathcal{H}$, where \mathcal{H} is a complex Hilbert space of arbitrary dimension. This means that the hidden variables are in one to one correspondence to pure states $\lambda \mapsto |\lambda\rangle$.

With this in mind, we can define the response functions and the transfer functions quantum mechanically for any instrument A and outcome $a \in O_A$:

$$\xi_A(a|\lambda_0) := \text{Tr} [\mathcal{I}_A^a (|\lambda_0\rangle \langle \lambda_0|)] \quad (\text{II.14})$$

$$\rho_a := \frac{\mathcal{I}_A^a (|\lambda_0\rangle \langle \lambda_0|)}{\xi_A(a|\lambda_0)} \quad (\text{II.15})$$

$$\Gamma_A(\lambda_1|\lambda_0, a) := \text{Tr} [\rho_a |\lambda_1\rangle \langle \lambda_1|] \quad (\text{II.16})$$

where \mathcal{I}_A^a here holds for the quantum instrument of A given outcome a , where we use the definition of [30] and [39, Chapter 4]. A quantum instrument is defined as a set of Completely Positive Linear Map $\mathcal{I}_A \equiv \{\mathcal{I}_A^a\}$ and which satisfy $\mathcal{I}_A^{*a}(\mathbb{1}) = M_A^a$ for every $a \in O_A$ where \mathcal{I}_A^{*a} is the dual map and M_A^a is quantum POVM element. The above equations can be generalized to mixed states knowing that the instruments are linear operators.

III. OPERATIONAL CONSIDERATIONS AND HVM RESTRICTIONS

Unrestricted HVMs can reproduce any theory, however there are some properties that we observe in experiments or that we believe to hold true and which constrains the behaviour of HVMs. The most seminal example is Bell theorem, where local causality [40, 41] is assumed to hold for any theory, thus it restricts the choice of HVMs to the ones respecting local causality. We can also operationally check it by doing the Bell experiment, and look at the resulting probability table, if however it is not respected then we can infer that local causality does not hold.

On the other hand, one can induce HVMs restrictions from operational considerations, similarly to what Spekkens proposed for his notion of contextuality [2]. In this case we assume that two preparations are equivalent if no measurement can distinguish them. Hence, if one makes an experiment with a tomographically complete set of measurements, and observe no difference in the statistics of two preparations, then we say the preparations are equivalent. As a result we can reflect this restriction on the HVM, by requiring the preparation functions to be the same.

In this section, we introduce three HVM restrictions: outcome determinism, no-disturbance and outcome independence, and we will relate them to operational considerations when we can. We relate these to non-contextuality in Section IV.

In an HVM, an instrument A will be characterized by its response function ξ_A and its transfer function Γ_A as defined in section II B.

A. Outcome determinism

We straightforwardly give the definition of outcome determinism:

Definition 8 (Outcome determinism). *In a HVM, an instrument A is called outcome deterministic if the probabilities of its classical outcome are always 0 or 1, more formally:*

$$\forall a \in O_A: \forall \lambda \in \Lambda: \xi_A(a|\lambda) \in \{0, 1\}. \quad (\text{III.1})$$

The assumption of quantum projectors may operationally justify outcome determinism [1, 18, 36, 42]. Assuming preparation non-contextuality can also motivate outcome determinism [2].

B. No-disturbance

An instrument A does not disturb another instrument B if placing A right before B does not change the probability of outcomes of B . As an example, a fair coin flip is never disturbed by any previous instrument, because it will always outcome either head or tails with half probability. Similarly, a non-invasive instrument will never disturb any following instrument, because since the system is not disturbed the following instrument must act on the undisturbed system. Any repeatable outcome deterministic instrument is also self non-disturbing. We analyse these examples after introducing the formal definition of no-disturbance for a HVM:

Definition 9 (No-disturbance). *In a HVM, an instrument A does not disturb another instrument A' , which we write $A \vdash A'$, if the following holds:*

$$\forall \lambda \in \Lambda: \forall a' \in O_{A'}: \sum_{a, \lambda'} \xi_A(a|\lambda) \Gamma_A(\lambda'|\lambda, a) \xi_{A'}(a'|\lambda') = \xi_{A'}(a'|\lambda) \quad (\text{III.2})$$

Meaning that the outcome statistics remain unchanged for A' after A has been measured.

Definition 9 is equivalent to definition 2 but reformulated for HVMs. Furthermore, it should be noted that Definition 9 is not necessarily symmetric, A' might not be disturbed by A , and yet A can be disturbed by A' .

Verifying Equation (III.2) requires one to be able to prepare every hidden variable $\lambda \in \Lambda$, which operationally means that one should have

a tomographically complete set of preparations. We point out that this assumption is commonly discussed in operational contextuality [43–45] and in the more general settings of GPT [46]. When one is convinced of operational no-disturbance, through many runs of the experiment for instance, then it is necessary to assume that there is no fine-tuning so that it also holds at the HVM level.

1. Example: Fair coin flip

The fair coin flip instrument A_{fcf} is never disturbed by any previous instrument. In order to prove it let us use Definition 9, where we have the fair coin flip response function to be $\xi_{A_{\text{fcf}}}(0|\lambda) = \xi_{A_{\text{fcf}}}(1|\lambda) = 0.5$ for all hidden variable $\lambda \in \Lambda$. Then:

$$\begin{aligned} & \sum_{a, \lambda'} \xi_A(a|\lambda) \Gamma_A(\lambda'|\lambda, a) \xi_{A_{\text{fcf}}}(a_{\text{fcf}}|\lambda') \\ &= \frac{1}{2} \sum_{a, \lambda'} \xi_A(a|\lambda) \Gamma_A(\lambda'|\lambda, a) \\ &= \frac{1}{2} \end{aligned}$$

where this holds for any instrument A and any $\lambda \in \Lambda$. Thus, the fair coin flip is never disturbed by any previous instrument, however it might disturb further instruments, depending on the specific state transformation that it implements.

2. Example: Non-invasive instrument

A non-invasive instrument A_{NI} , is a device that does not change the hidden variable, i.e., $\Gamma_{A_{\text{NI}}}(\lambda'|\lambda) = \delta_{\lambda, \lambda'}$. In that case, for any following instrument A Definition 9 is satisfied:

$$\begin{aligned} & \sum_{a_{\text{NI}}, \lambda'} \xi_{A_{\text{NI}}}(a_{\text{NI}}|\lambda) \Gamma_{A_{\text{NI}}}(\lambda'|\lambda, a) \xi_A(a|\lambda') \\ &= \sum_{a_{\text{NI}}, \lambda'} \xi_{A_{\text{NI}}}(a_{\text{NI}}|\lambda) \delta_{\lambda, \lambda'} \xi_A(a|\lambda') \\ &= \xi_A(a|\lambda) \end{aligned}$$

Thus, it is quite clear that a non-invasive instrument is always non-disturbing according to our definition.

3. Example: Projective deterministic instrument

We define a repeatable deterministic instrument A_{rep} as a device that has a deterministic response function and once put in sequence always output the same classical outcome. Mathematically it means that for any hidden variable $\lambda \in \Lambda$ and any classical outcome $a_{\text{rep}} \in O_{A_{\text{rep}}}$ we must have these two conditions:

$$\xi_{A_{\text{r}}}(a_{\text{r}}|\lambda) \in \{0, 1\} \quad (\text{III.3})$$

$$\sum_{\lambda'} \xi_{A_{\text{r}}}(a_{\text{r}}|\lambda) \Gamma_{A_{\text{r}}}(\lambda'|\lambda, a_{\text{r}}) \xi_{A_{\text{r}}}(a'_{\text{r}} = a_{\text{r}}|\lambda') = 1 \quad (\text{III.4})$$

Using equation (III.4) we see that whenever $\xi_{A_{\text{r}}}(a_{\text{r}}|\lambda) = 1$ it must be that $\xi_{A_{\text{r}}}(a'_{\text{r}} \neq a_{\text{r}}|\lambda') = 0$ on the support of $\Gamma_{A_{\text{r}}}$. Injecting into equation (III.2) we obtain:

$$\begin{aligned} & \sum_{a_{\text{r}}, \lambda'} \xi_{A_{\text{r}}}(a_{\text{r}}|\lambda) \Gamma_{A_{\text{r}}}(\lambda'|\lambda, a_{\text{r}}) \xi_{A_{\text{r}}}(a'_{\text{r}}|\lambda') \\ &= \sum_{a_{\text{r}}, \lambda'} \xi_{A_{\text{r}}}(a_{\text{r}}|\lambda) \Gamma_{A_{\text{r}}}(\lambda'|\lambda, a_{\text{r}}) \xi_{A_{\text{r}}}(a'_{\text{r}} = a_{\text{r}}|\lambda') \\ &= \xi_{A_{\text{r}}}(a_{\text{r}}|\lambda) \end{aligned} \quad (\text{III.5})$$

Which holds for all $\lambda \in \Lambda$ and $a'_{\text{r}} \in O_{A_{\text{r}}}$.

4. No-disturbance in quantum theory

The notion of no-disturbance is defined for POVM measurements in quantum theory in [30]. We provide an adapted definition for instruments below:

Definition 10 ((Quantum) Non-disturbing instruments from [30]). *Given two instruments A and B , A can be performed without disturbing B , if:*

$$\forall \rho \in \mathcal{T}(\mathcal{H}) : \forall b \in O_B : \quad \text{Tr} \left[\mathcal{I}_B^b \left(\sum_a \mathcal{I}_A^a(\rho) \right) \right] = \text{Tr} [\mathcal{I}_B^b(\rho)] \quad (\text{III.6})$$

Where \mathcal{H} is a Hilbert space of arbitrary dimension and $\mathcal{T}(\mathcal{H})$ is the set of positive trace class operators of trace one.

We observe that HVMs adapted to quantum theory as shown by Section II C (equations (II.14)–(II.16)) in fact retrieve equation (III.6) when applied to our definition of no-disturbance (see Definition 9).

C. Outcome independence

We also introduce another potential restriction called *outcome independence* (OI). The transfer function Γ is OI if it only depends on the previous hidden variable λ and not on the previous outcome a . Namely, the new hidden variable λ' after the measurement must be independent of the outcome a .

Such a typical instrument is an instrument that samples randomly after the classical outcome production, or the opposite, an instrument that deterministically produces the same hidden variable after the classical outcome. We will go more in depth with these examples, but first we provide the definition of outcome independence:

Definition 11 ((Transfer functions) Outcome independence). *In a HVM, the transfer function Γ_A of an instrument A is outcome independent, if for all outcome, the probability of obtaining the state λ' does not change:*

$$\forall \lambda, \lambda' \in \Lambda: \forall a, a' \in O_A: \Gamma_A(\lambda'|\lambda, a) = \Gamma_A(\lambda'|\lambda, a'). \quad (\text{III.7})$$

For simplicity, we will drop the dependency in a to denote a transfer function that is outcome independent, i.e., $\Gamma_A(\lambda'|\lambda) \equiv \Gamma_A(\lambda'|\lambda, a)$.

We would like to highlight the fact that outcome independence is difficult to operationally motivate, particularly since quantum mechanics does have outcome dependence. It is not a requirement to show contextuality, however one can retrieve a notion of contextuality from this assumption, which we make explicit in Theorem 4.

1. Example : Random resampling

An instrument A_{rr} is randomly resampling if given a set of hidden variables Λ we have:

$$\Gamma_{A_{\text{rr}}}(\lambda'|\lambda, a_{\text{rr}}) = \frac{1}{|\Lambda|} \quad (\text{III.8})$$

Clearly, for such an instrument, the transfer function does not depend on the classical outcome a_{rr} .

2. Example : Deterministic reset

Similarly to the random resampling, a deterministic reset instrument A_{dr} , is an instrument

that resets the hidden variable λ after producing the classical outcome a_{dr} :

$$\Gamma_{A_{\text{dr}}}(\lambda'|\lambda, a_{\text{dr}}) = \delta_{\lambda_{\text{dr}}, \lambda'} \quad (\text{III.9})$$

where λ_{dr} is a given hidden variable. Thus, the instrument simply always outputs the same hidden variable, whatever the input, and again, this does not depend on the classical outcome.

3. Programmable instruments and OI

The notion of outcome independence has already appeared in the work of Buscemi et al. [29], where they introduced a specific type of programmable instrument – essentially an instrument with two inputs and outputs, one classical and one quantum – where the production of the classical output is detached from the production of the quantum output, without the possibility to communicate. It is interesting to notice that some type of compatibility, termed “classical compatibility” [29, Def. 2], is inherently OI.

IV. NON-CONTEXTUALITY FOR SEQUENTIAL SCENARIOS

We now show how the restrictions introduced in Section III relate to non-contextuality, specifically as defined in the sheaf theoretic framework. We start with the definition of non-contextuality, and we show that non-contextual empirical models form a polytope. Not only that, but we discuss extensively on the contextual fraction, what it means in sequential scenarios, and we frame some related open questions.

First, we define the unique set of instruments in a sequence S as $\bar{S} = \{A \mid \exists i \in \mathbb{N}, A_i \in S\}$, this is basically the set of all instruments in a sequence where we removed the indices. As an example, if $S = \{A_1, B_2, A_3, C_4, B_5\}$ then we have $\bar{S} = \{A, B, C\}$. We also define a consistent outcome:

Definition 12 (Consistent outcome). *An outcome is said to be consistent, if within a sequence $S \in \mathcal{S}$ the outcome of an instrument is the same regardless of its position in the sequence S . More precisely, in a scenario $\langle X, \mathcal{S}, O \rangle_{\text{seq}}$ we say that an outcome $o \in O_S$, with $S \in \mathcal{S}$, is consistent if there exists $o' \in O_{\bar{S}}$ such that:*

$$\forall A \in \bar{S}: \forall i \in \mathbb{N}: A_i \in S \implies o|_{A_i} = o'|_A. \quad (\text{IV.1})$$

If such an o' exist for o we will denote it by $o_{\bar{S}}$.

We directly show that $o_{\bar{S}}$ is unique:

Lemma 1 (Uniqueness of $o_{\bar{S}}$). $o_{\bar{S}} \in O_{\bar{S}}$ as in Definition 12 is unique.

Proof. We prove it by contradiction. Let $o_{\bar{S}}, o'_{\bar{S}} \in O_{\bar{S}}$ such that $o_{\bar{S}} \neq o'_{\bar{S}}$ and they both satisfy Equation (IV.1). Necessarily we have for all $A \in \bar{S}$:

$$o_{\bar{S}}|_A = o'_{\bar{S}}|_A \quad (\text{IV.2})$$

Thus they must be the same, which contradicts $o_{\bar{S}} \neq o'_{\bar{S}}$. \square

We now define non-contextuality for sequential scenarios, following closely the definitions in [3], but where we allow for repeating instruments, i.e., an instrument may appear multiple times in the same sequence.

Definition 13 (Non-contextuality for sequential scenarios). *We say that an empirical model e on a sequential scenario $\langle X, \mathcal{S}, O \rangle_{\text{seq}}$ is non-contextual, if there exist a global probability distribution on outcomes $d: O_X \rightarrow [0, 1]$, such that any marginal in the empirical model can be retrieved as a marginal of the global probability distribution d if the outcomes are consistent, i.e., a repeated instrument in a sequence has the same outcome. Then for all $S \in \mathcal{S}$ and for all $o \in O_S$:*

$$e_S(o) = \begin{cases} d|_{\bar{S}}(o_{\bar{S}}), & \text{if } o \text{ is consistent} \\ 0, & \text{otherwise} \end{cases} \quad (\text{IV.3})$$

Remark : For any sequence $S \in \mathcal{S}$ if there exists another sequence $S' \in \mathcal{S}$ such that $\bar{S} = \bar{S}'$ then they must have the same probability distribution over consistent outcomes. This can be directly inferred from Equation (IV.3).

A. Convexity of non-contextual empirical models

One natural question is to wonder whether the set of non-contextual behaviour is convex, and whether the convex set can be defined as a polytope or not. We prove that it is in fact a convex polytope which means that a mixture of non-contextual empirical models must yield a non-contextual empirical model. We argue that the extremal point of the non-contextual polytope for sequential scenarios are also given by the global deterministic assignments on O_X .

Theorem 1. *The set of non-contextual behaviours form a polytope, where the extremal points are given by global deterministic assignments.*

Proof. By definition, we say that an empirical model e on a sequential scenario $\langle X, \mathcal{S}, O \rangle_{\text{seq}}$ is non-contextual if there exist a global distribution on the outcomes $d: O_X \rightarrow [0, 1]$ and it respects Equation IV.3. First we can see that such a global probability distribution can be seen as a mixture of deterministic global assignments:

$$d = \sum_i \mu_i d^{\text{det}_i} \quad (\text{IV.4})$$

where $d^{\text{det}_i}: O_X \rightarrow \{0, 1\}$ is a deterministic assignment of outcomes and $\{\mu_i\}_i$ gives a probability distribution over these deterministic assignments. Each of these d^{det_i} give rise to a deterministic empirical model given by, for all $S \in \mathcal{S}$:

$$e_S^{\text{det}_i}(o) = \begin{cases} d^{\text{det}_i}|_{\bar{S}}(o_{\bar{S}}), & \text{if } o \text{ is consistent} \\ 0, & \text{otherwise} \end{cases} \quad (\text{IV.5})$$

Then we can rewrite Equation (IV.3) as the following:

$$e_S(o) = \begin{cases} \sum_i \mu_i d^{\text{det}_i}|_{\bar{S}}(o_{\bar{S}}), & \text{if } o \text{ is consistent} \\ 0, & \text{otherwise} \end{cases} \quad (\text{IV.6})$$

More concisely, this may be written as:

$$e_S(o) = \sum_i \mu_i e_S^{\text{det}_i}(o) \quad (\text{IV.7})$$

Thus we have just proven that any non-contextual model can be written as a convex mixture of deterministic non-contextual empirical model, which are the vertices of the polytope. \square

Because sequential scenarios allow for repeated instruments inside a context, we can wonder whether it impacts the shape of the polytope to add a repeated instrument at the end of a context. For instance, let us consider the following extended sequential KCBS scenario:

Example 2 (Extended sequential KCBS). *Let the sequential KCBS scenario as per Example 1, where we take a different sequence S :*

$$\mathcal{S}_{E\text{-KCBS}} = \{ \{A_1^0, A_2^1, A_3^0\}, \{A_1^1, A_2^2\}, \{A_1^2, A_2^3\}, \{A_1^3, A_2^4\}, \{A_1^4, A_2^0\} \} \quad (\text{IV.8})$$

where we have simply added A_3^0 at the end of the first sequence.

By construction, the spaces of empirical models, or correlations, of Example 1 and Example 2 are different (since more outcomes are involved in Example 2). However, since the set of instruments is the same, by Theorem 1 the non-contextual polytopes have the same number of extremal points, that are given by the fixed assignments for each instrument.

B. Contextual fraction for sequential scenarios

Once we have an empirical model e on a given sequential scenario $\langle X, \mathcal{S}, O \rangle_{\text{seq}}$, we can require not only to have a qualitative measure of contextuality but rather a quantitative measure of contextuality, i.e., how contextual is a given empirical model. In the standard sheaf theoretic contextuality scenario, contextuality can be quantified with the contextual fraction [3], and this can be implemented as a linear program [47]. Thus, we naturally extend the definition of the contextual fraction for sequential scenarios in this section.

In order to define the contextual fraction, we first start by decomposing any sequential empirical model as a convex sum of empirical models:

$$e = \lambda e^{\text{NC}} + (1 - \lambda)e' \quad (\text{IV.9})$$

where e^{NC} is a non-contextual model as per Definition 13 and e' is any empirical model. If one maximizes the weight on λ , which takes the optimal value λ^* , then we can define the non-contextual fraction of e as $\text{NCF}(e) := \lambda^*$, and by definition the contextual fraction is defined as $\text{CF}(e) := 1 - \text{NCF}(e)$. The contextual fraction was originally designed for empirical models on measurement scenarios (See Section VI), but we see from the definition that it naturally extends to empirical models on sequential scenarios. It is a convex measure [47] and it is continuous [48].

In section IV A we questioned whether addition of repeated measurements changes the NC polytope. Here we extend the question to quantum behaviour produced by quantum instruments and states, where we question the difference that may arise from repeating measurements in sequences and if it has an impact on the contextual fraction. More precisely, let a quantum realization of the sequential scenario given by Example 1, and in

particular let us take the common quantum realization made by projective measurements, see references [36, 37] for the exact set, and Lüders instruments. Then by keeping the same measurements and instruments but using the sequential scenario of the extended KCBS, i.e. Example 2, it is not clear how the contextual fraction will evolve. For that specific example, i.e. sequential KCBS vs extended sequential KCBS, we have numerically shown that it does not increase.

Open question 1. *In the general case, it may be the case that repeating measurements inside a sequential scenario give more or less non-contextuality. Answering this question would lead to a better understanding of sequential scenarios, and it is left open for future work.*

V. RESTRICTED HVMS AND NON-CONTEXTUALITY

In the previous section we have introduced the definition of contextuality, but it is of interest to derive this notion of contextuality from restricted HVM. In this section we make the formal link between restricted HVMS and non-contextual HVMS.

First we define what is a non-disturbing hidden variable model (ND HVM):

Definition 14 (ND HVM). *A HVM $\langle \Lambda, \mu, (h^\lambda)_{\lambda \in \Lambda} \rangle$ on a sequential scenario $\langle X, \mathcal{S}, O \rangle_{\text{seq}}$ is non-disturbing, if for any sequence $S \in \mathcal{S}$ and any two instruments in that sequence $A_i \neq A_j \in S$ we have:*

$$j > i \implies A_i \vdash A_j. \quad (\text{V.1})$$

Where the notation $A_i \vdash A_j$ means that A_i does not disturb A_j . We will refer to any behaviour of such a hidden variable model as an ND HVM model.

This can be understood as the property that for all sequences, any instrument in it is such that it will not disturb any other instrument later.

We refer to an outcome deterministic HVM $\langle \Lambda, \mu, (h^\lambda)_{\lambda \in \Lambda} \rangle$ on a sequential scenario $\langle X, \mathcal{S}, O \rangle_{\text{seq}}$, whenever every instrument is outcome deterministic in the sense of Definition 8:

Definition 15 (OD HVM). *A HVM $\langle \Lambda, \mu, (h^\lambda)_{\lambda \in \Lambda} \rangle$ on a sequential scenario $\langle X, \mathcal{S}, O \rangle_{\text{seq}}$ is outcome deterministic if for*

any sequence $S \in \mathcal{S}$, all instruments in such a sequence $A_i \in \mathcal{S}$ are OD as in Definition 8.

Finally, we talk of an outcome independent (OI) HVM:

Definition 16 (OI HVM). A HVM $\langle \Lambda, \mu, (h^\lambda)_{\lambda \in \Lambda} \rangle$ on a sequential scenario $\langle X, \mathcal{S}, O \rangle_{\text{seq}}$ is outcome independent if for any sequence $S \in \mathcal{S}$, all instruments in such a sequence $A_i \in \mathcal{S}$ are OI as in Definition 11.

A. No-disturbance and the compatibility of marginals

In the sheaf theoretic framework for contextuality, if there exists a HVM which respects both the compatibility of marginals (generalization of no-signalling) and outcome determinism then it is equivalent to non-contextuality. In this paper we have been looking at no-disturbance, which do imply compatibility of marginals, in the addition of no-signalling from the future. More formally:

Theorem 2 (ND implies the compatibility of marginals). Let a ND HVM behaviour h on a sequential scenario $\langle X, \mathcal{S}, O \rangle_{\text{seq}}$, if h respects the NBTS condition then it also respects the compatibility of marginals.

For the full proof see Appendix A 1, the general idea is to use ND as one way no signalling, and causality as the other way no-signalling. With these assumptions, one ends up by respecting the compatibility of marginals.

As a reminder, the compatibility of marginals (see Definition 5) is operational, but it is a mathematical property rather than a physical observation. The notion of no-disturbance is on the other hand operational and physical, in the same way that one uses space like separations in Bell experiments to justify the compatibility of marginals. Thus, these two notions are not the same, but one implies the other, as Theorem 2 attests.

B. No-disturbance with outcome determinism is equivalent to non-contextuality for sequential scenarios

After Theorem 2 shows us that any ND HVM on a sequential scenario implies the compatibility of marginals we now turn to OD ND HVM, and we show that these hidden variables are non-contextual.

Theorem 3 (ND and OD HVM are equivalent to non-contextuality). Let an empirical model e on a sequential scenario $\langle X, \mathcal{S}, O \rangle_{\text{seq}}$. This empirical model is non-contextual if and only if there exists a ND OD HVM $\langle \Lambda, \mu, (h^\lambda)_{\lambda \in \Lambda} \rangle$ respecting the NBTS condition which realizes it.

The proof is developed in appendix A 2.

We now turn to seminal examples in the literature for contextuality, the KCBS inequality [36] for which we gave the shape of the scenario in Example 1 and the Peres-Mermin square [49, 50]. We show below that both examples can still witness contextuality in sequential scenarios. In both examples below we will make use of Lüders instruments [51]: given a POVM $E^x = \{E^{x,y}\}_{y \in O_{E^x}}$ its Lüders instrument is defined for all $y \in O_{E^x}$ as $\mathcal{I}_{E^x}^y(\rho) = \sqrt{E^{x,y}}\rho\sqrt{E^{x,y}}$.

The derivation of the inequalities below can be done following the same reasoning as in standard contextuality scenarios. We formalize this general connection in section VI with Theorem 5. These examples now allow us to interpret these inequalities in the sequential scenario versions.

Example 3 (KCBS non-contextuality in sequential scenarios). We start from Example 1 with the sequential scenario $\langle X_{\text{KCBS}}, \mathcal{S}_{\text{KCBS}}, O_{\text{KCBS}} \rangle_{\text{seq}}$. Suppose that you can operationally justify determinism and non disturbance, i.e., for all $i \in \{0, \dots, 4\}$ we have $A^i \vdash A^{i \oplus 1}$. Then the following inequality [6, 36] holds by virtue of Theorem 3:

$$\sum_{i=0}^4 p(a^i = a^{i \oplus 1} \mid A_0^i A_1^{i \oplus 1}) \geq 1 \quad (\text{V.2})$$

Where $i \oplus 1$ is the addition modulo 5. In order to witness contextuality, one can follow the guideline of quantum theory to derive an experimental setup. It is important to stress that while the quantum formalism can be used as a tool to derive an experimental setup, the assumptions are operational.

Let M^{A^i} be a measurement with binary outcomes $M^{A^i} = \{P_j^{A^i}\}_{j=0,1}$ where $P_j^{A^i}$ is a projector that correspond to outcome j . Their associated instrument is $\mathcal{I}^{A^i} = \{\mathcal{I}_j^{A^i}\}_{j=0,1}$ such that for all $i \in \{0, \dots, 4\}$ and for all $j \in \{0, 1\}$ we have $\mathcal{I}_j^{A^i}(\rho) = P_j^{A^i} \rho P_j^{A^i}$. We will take one of the

quantum realization of the KCBS inequality [52]:

$$\begin{aligned} |\nu_0\rangle &= \frac{1}{N} \begin{pmatrix} 1 \\ 0 \\ \sqrt{\cos\theta} \end{pmatrix} & |\nu_1\rangle &= \frac{1}{N} \begin{pmatrix} \cos(4\theta) \\ \sin(4\theta) \\ \sqrt{\cos\theta} \end{pmatrix} \\ |\nu_2\rangle &= \frac{1}{N} \begin{pmatrix} \cos(2\theta) \\ -\sin(2\theta) \\ \sqrt{\cos\theta} \end{pmatrix} & |\nu_3\rangle &= \frac{1}{N} \begin{pmatrix} \cos(2\theta) \\ \sin(2\theta) \\ \sqrt{\cos\theta} \end{pmatrix} \\ |\nu_4\rangle &= \frac{1}{N} \begin{pmatrix} \cos(4\theta) \\ -\sin(4\theta) \\ \sqrt{\cos\theta} \end{pmatrix}. \end{aligned}$$

Where $\theta = \pi/5$, $N = \sqrt{1 + \cos\theta}$ and:

$$\begin{aligned} P_0^{A^i} &= |\nu_i\rangle\langle\nu_i| \\ P_1^{A^i} &= I - |\nu_i\rangle\langle\nu_i|. \end{aligned}$$

Where I is the identity for a qutrit.

Since we have pairwise commutation for all $i \in \{0, \dots, 4\}$:

$$\forall j \in \{0, 1\}: [P_j^{A^i}, P_j^{A^{i\oplus 1}}] = 0.$$

Then instrument no-disturbance must hold [30] (see Definition 10):

$$\forall i \in \{0, \dots, 4\}: \mathcal{I}^{A^i} \vdash \mathcal{I}^{A^{i\oplus 1}}.$$

Yet for the state $\langle\psi| = (0, 0, 1)$, we observe a violation of inequality (V.2):

$$\sum_{i=0}^4 p(a^i = a^{i\oplus 1} | A_0^i A_1^{i\oplus 1}) = 5 - 2\sqrt{5} \not\geq 1.$$

Example 4 (Peres-Mermin non-contextuality in sequential scenarios). *We show that the Peres-Mermin Square [7, 49, 50] can be encompassed in this framework if it is done sequentially. First, we start by ordering the instruments in the contexts, to give sequences, which is also done in Ref [53] for example. Thus, the sets in $\langle X_{\text{PM}}, \mathcal{S}_{\text{PM}}, O_{\text{PM}} \rangle_{\text{seq}}$ are:*

$$X_{\text{PM}} = \{A^i\}_{1 \leq i \leq 9} \quad (\text{V.3})$$

$$\begin{aligned} \mathcal{S}_{\text{PM}} &= \left\{ \{A_0^i, A_1^{i+1}, A_2^{i+2}\} \right\}_{i \in \{1, 4, 7\}} \\ &\cup \left\{ \{A_0^i, A_1^{i+3}, A_2^{i+6}\} \right\}_{i \in \{1, 2, 3\}} \end{aligned} \quad (\text{V.4})$$

$$\forall x \in X: O_{\text{PM}}^x = \{0, 1\} \quad (\text{V.5})$$

Where, similarly to the KCBS Example 1, we use indices for the order and exponents for the instrument. Again if one assumes determinism and no-disturbance the following inequality holds:

$$\sum_{S \in \mathcal{S}_{\text{PM}}} p(a^i a^j a^k = (-1)^{\delta_{i,3}\delta_{j,6}\delta_{k,9}} | A_0^i A_1^j A_2^k) \leq 5. \quad (\text{V.6})$$

Where $A_0^i A_1^j A_2^k = S \in \mathcal{S}_{\text{PM}}$. The PM square inequality can be violated in quantum theory with the observables $O^{A^i} = P_0^{A^i} - P_1^{A^i}$ such that the instruments \mathcal{I}^{A^i} are defined in the same way as Example 3. The set of observables is the following:

$$\begin{aligned} O^{A^1} &= \sigma_z \otimes \mathbb{1} & O^{A^2} &= \mathbb{1} \otimes \sigma_z & O^{A^3} &= \sigma_z \otimes \sigma_z \\ O^{A^4} &= \mathbb{1} \otimes \sigma_x & O^{A^5} &= \sigma_x \otimes \mathbb{1} & O^{A^6} &= \sigma_x \otimes \sigma_x \\ O^{A^7} &= \sigma_z \otimes \sigma_x & O^{A^8} &= \sigma_x \otimes \sigma_z & O^{A^9} &= \sigma_y \otimes \sigma_y. \end{aligned}$$

Where $\sigma_{x,y,z}$ are the Pauli matrices and $\mathbb{1}$ is the identity.

Similarly to the previous example, we see that all measurements are projective and commuting within a sequence $S \in \mathcal{S}$. The associated quantum instruments must therefore be non-disturbing:

$$\forall S \in \mathcal{S}_{\text{PM}}: \forall A_i^k, A_j^\ell \in S: i > j: \mathcal{I}^{A^k} \vdash \mathcal{I}^{A^\ell}.$$

Finally, from the quantum example we get the following violation of Inequality (V.6):

$$\sum_{S \in \mathcal{S}_{\text{PM}}} p(a^i a^j a^k = (-1)^{\delta_{i,3}\delta_{j,6}\delta_{k,9}} | A_0^i A_1^j A_2^k) = 6 \not\leq 5$$

C. No-disturbance with outcome independence is equivalent to non-contextuality for sequential scenarios

In this section we show that in a sequential scenario, an empirical model is non-contextual if there exists a ND OI HVM which realizes it, as long as we restrict ourselves to sequences $S \in \mathcal{S}$ of 2 instruments or less. Restricting to sequences of length less than 2 still covers numerous interesting contextuality scenarios, such as all the n-cycle contextuality scenarios [54]. We formalize it with the following theorem:

Theorem 4 (ND and OI HVM are equivalent to non-contextuality). *Let an empirical model e on a sequential scenario $\langle X, \mathcal{S}, O \rangle_{\text{seq}}$ such that:*

$$\forall S \in \mathcal{S}: |S| \leq 2 \quad (\text{V.7})$$

This empirical model is non-contextual if and only if there exists a ND OI HVM $\langle \Lambda, \mu, (h^\lambda)_{\lambda \in \Lambda} \rangle$ respecting the NBTS condition which realizes it.

The full proof is given in the Appendix A 3.

VI. MEASUREMENT SCENARIO TO SEQUENTIAL SCENARIOS

One of the motivations for studying no-disturbance in sequential scenarios is its operational significance. In this spirit, it is reasonable to ask whether any measurement scenario as described by the sheaf theoretic framework can be transposed in a sequential scenario. We answer by the positive constructively. We moreover show that, for our construction, the non-contextual polytope is conserved, thus the non-contextual inequalities are preserved through what we call the induced sequential scenario.

Open question 2. *The question for the quantum set, i.e., whether it holds in the induced sequential scenarios is left open. We strongly believe that it should not be the same, since we are allowing any type of instruments, but at this time we are not able to provide a proof of this claim.*

In the following, we first make a quick introduction to measurement scenarios and empirical scenarios in the sheaf theoretic framework for contextuality, then we show how one can map any measurement scenario to a sequential scenario.

A. Measurement scenario à la sheaf theoretic framework

We restrict ourselves to an informal description of the sheaf theoretic framework, with enough details for the reader to follow. A more thorough description can be found in the original paper [3] and in subsequent papers [47, 48].

A *measurement scenario* is characterized by the tuple $\langle X, \mathcal{M}, O \rangle$ where (i) X is a set of *measurement labels*; (ii) $\mathcal{M} \subseteq \mathcal{P}(X)$ is the set of maximal contexts and forms a cover of X ; (iii) $O = (O_x)_{x \in X}$ and O_x is the outcome set corresponding to measurement $x \in X$.

One can define an *empirical model* e on a measurement scenario $\langle X, \mathcal{M}, O \rangle$ as a family of probability distributions $e = (e_C)_{C \in \mathcal{M}}$, in the same spirit as an empirical model on a sequential scenario (see Definition 4).

The compatibility of marginals, referring to the operational property of no-signalling, parameter independence or “no-disturbance”, is substantially simplified in measurement scenarios. An empirical model is said to respect the compatibility of marginals if for any two contexts C_1 and C_2 we have, $e_{C_1}|_{C_1 \cap C_2} = e_{C_2}|_{C_1 \cap C_2}$ where the notation $e_C|_U$ stands for the marginalization of the probability distribution e_C to $U \subseteq C$.

Non-contextuality is the fact that an empirical model e on a measurement scenario $\langle X, \mathcal{M}, O \rangle$ can be described by a global probability distribution on the outcomes $d: O_X \rightarrow [0, 1]$, i.e., all context wise probability distributions $e_C: O_C \rightarrow [0, 1]$ can be seen as the marginal of d , or, mathematically, $e_C = d|_C$.

We follow up with the definition of a *hidden variable model*. A HVM on a measurement scenario $\langle X, \mathcal{M}, O \rangle$ is a triple $\langle \Lambda, \mu, (h^\lambda)_{\lambda \in \Lambda} \rangle$, where (i) Λ is a finite set of hidden variables; (ii) $\mu: \Lambda \rightarrow [0, 1]$ is the distribution over hidden variables; (iii) for all λ we have h^λ as an empirical model on $\langle X, \mathcal{M}, O \rangle$. This gives rise to an overall empirical model given as the average over the various hidden variables $h := \sum_{\lambda \in \Lambda} \mu(\lambda) h^\lambda$. Similarly to empirical models, we can impose that a given hidden variable behaviour h^λ must be parameter independent, i.e. behaviours which respect the compatibility of marginals, such that for all $C_1, C_2 \in \mathcal{M}$ we have $h_{C_1}^\lambda|_{C_1 \cap C_2} = h_{C_2}^\lambda|_{C_1 \cap C_2}$. Imposing such a constraint on each of the hidden variable behaviours lead to an empirical model which respect the compatibility of marginals, such that for all $C_1, C_2 \in \mathcal{M}$:

$$\begin{aligned} \forall \lambda \in \Lambda: h_{C_1}^\lambda|_{C_1 \cap C_2} &= h_{C_2}^\lambda|_{C_1 \cap C_2} \\ \implies h_{C_1}|_{C_1 \cap C_2} &= h_{C_2}|_{C_1 \cap C_2} \end{aligned} \quad (\text{VI.1})$$

Finally, we can define a deterministic, or sharp, hidden variable behaviour, such that $\forall \lambda: \forall C \in \mathcal{M}: h_C^\lambda(t) = \delta_{t,t'}$ where $t, t' \in O_C$ and δ is the Dirac distribution.

We write below the conclusion reached in [3, Prop. 3.1 and Th. 8.1]:

Proposition 1. *An empirical behaviour e is non-contextual if and only if it is realizable by a parameter independent deterministic HVM.*

B. Mapping measurement scenarios to sequential scenarios

From the definitions given in the previous section, we are able to show that any measurement

scenario can be converted to a sequential scenario, where the non-contextual set is preserved. This means the notion of no-signalling is replaced by no-disturbance and measurements are turned into instruments.

Let an arbitrary measurement scenario $\langle X, \mathcal{M}, O \rangle$, with measurement labels $X = \{A_i\}_{i \in I}$, without loss of generality the contexts are assumed to take the form $C_j = \{A_k\}_{k \in K \subseteq I} \in \mathcal{M}$. We define the number of contexts to be $N = |\mathcal{M}|$, such that we can index them C_j with $j = 1, \dots, N$. Finally, each measurement has its own outcome set $(O_x)_{x \in X}$, which for simplicity we will assume to be the same for each measurement, i.e., $\forall x \in X: O_x = O$.

From such a measurement scenario, we can define a sequential scenario where each measurement is promoted to an instrument, i.e., they also output a system. In essence, this does not change the scenario, because a scenario is only concerned about the labelling of such measurements (or instruments). Then, one decides of an ordering for the contexts, in other words for each context C_j the instruments inside $A_k^\ell \in C_j$ acquire a new exponent ℓ representing the order of the instrument in the sequence. We call this new ordered context of instruments a sequence, which we write S_j . The choice of the ordering may be made arbitrarily, or decided such that the condition of no-disturbance is respected. With these changes, one can define a sequential scenario $\langle X_s, \mathcal{S}_s, O_s \rangle_{\text{seq}}$, with $X_s = X$, $\mathcal{S}_s = \{S_i\}_{i=1 \dots N}$ and the classical outcomes remain unchanged $O_s = O$. We call these sequential scenarios *induced sequential scenarios*:

Definition 17 (Induced sequential scenario). *A sequential scenario $\langle X_s, \mathcal{S}_s, O_s \rangle_{\text{seq}}$ is called an induced sequential scenario if there exist a measurement scenario $\langle X, \mathcal{M}, O \rangle$ such that (i) $X_s = X$; (ii) for all $S \in \mathcal{S}_s$, there exists $C \in \mathcal{M}$ such that $S \approx C$ up to the index of the instruments; (iii) for all $S \in \mathcal{S}_s$, there exists $C \in \mathcal{M}$ such that $O_C = O_S$.*

For a given measurement scenario, there exists multiple possible induced sequential scenario, where the various ordering of the instruments make the difference. However, there exists only one measurement scenario for a given induced scenario, which basically boils down to the fact that the contexts must be the same as the sequential scenario up to the ordering labelling.

We now prove that such an induced sequential scenario preserves the notion of non-contextuality: if an empirical model is non-contextual in the measurement scenario then it is non-contextual in the sequential scenario and vice versa. This implies that the existence of a deterministic parameter independent hidden variable model in the measurement scenario is equivalent to the existence of a deterministic non-disturbing hidden variable model in the induced sequential scenario.

C. Non-contextuality holds in induced sequential scenarios

Given this mapping, it is natural to expect that the non-contextual sets coincide. This is formalised below.

Theorem 5. *Let an empirical model e on a measurement scenario $\langle X, \mathcal{M}, O \rangle$ and an induced sequential scenario $\langle X, \mathcal{S}, O \rangle_{\text{seq}}$ built from it as described in Section VI B. Then if this empirical model e is NC in the measurement scenario, then it is NC in the induced sequential scenario and vice versa.*

Proof. We start by the first implication, that NC in measurement scenarios implies NC in induced sequential scenarios. Let a NC empirical model e on the measurement scenario $\langle X, \mathcal{M}, O \rangle$. Then there exists a global distribution $d: O_X \rightarrow [0, 1]$ such that for each context, the empirical scenario is retrieved as the marginal of that global probability distribution $e_C = d|_C$.

We define an analogous empirical model e_s on the sequential scenario $\langle X_s, \mathcal{S}_s, O_s \rangle_{\text{seq}}$ as the same empirical model, i.e. it is the same family of probability distributions, $e_s := e$ where the only modification in the contexts is the addition of a label to the measurements $A \in C \mapsto A_\ell \in S$. The following is valid for any labelling, as long as the label ℓ appears only once in a given sequence, and $\ell \leq |C|$. In order to show that the empirical model e_s satisfies the notion of non-contextuality described in Definition 13, we first note that every sequence S is the same as the set of unordered measurements: $S = \bar{S}$. This holds by definition of our mapping. Secondly, for any sequence S , the outcomes $o \in S$ are consistent as in Definition 12, since we have the property $S = \bar{S}^5$.

⁵ $S = \bar{S}$ implies that $O_S = O_{\bar{S}}$ and thus $\forall o \in O_S$ there

Because we have that any $o \in O_S$ is consistent, we can rewrite Equation (IV.3) as:

$$\forall S \in \mathcal{S}: \forall: e_S(o) = d_s|_{\bar{S}}(o_{\bar{S}}) \quad (\text{VI.2})$$

which holds for all $S \in \mathcal{S}$ and all $o \in O_S$. One can see that $\bar{S} = C$, such that if we take $d_s \equiv d$ we indeed retrieve Equation (IV.3).

For the other implication, that a NC empirical model in an induced sequential scenario give rise to a NC empirical model in a measurement scenario, the proof follows the same structure. Let a non-contextual empirical model e_s in a sequential scenario $\langle X, \mathcal{S}, O \rangle_{\text{seq}}$. Then there must exist a global probability distribution on the outcomes $d_s: O_X \rightarrow [0, 1]$ such that $e_S(o) = d|_{\bar{S}}(o_{\bar{S}})$. Note here that we use the fact previously shown that all outcomes are consistent and $\forall S \in \mathcal{S}: S = \bar{S}$. Thus, setting a global distribution in measurement scenario $d := d_s$, we find that the empirical model in the measurement scenario $e := e_s$ is also NC, i.e., for all $S \in \mathcal{S}$ and for all $o \in O_S$:

$$\begin{aligned} e_S(o) &= d_s|_{\bar{S}}(o_{\bar{S}}) \\ &= d|_C(o_C) \\ &= e_C(o) \end{aligned} \quad (\text{VI.3})$$

We thus conclude that the empirical model e on the measurement scenario must be non-contextual if e_s is non-contextual. \square

Although not surprising, this Theorem has important consequences, as it allows any contextual scenario in the standard framework to be translated, along with the associated inequalities, measures of contextuality (the contextual fraction) and tools that come with them. In particular, using [47] one can take experimental data from any sequential version of a contextual inequality and check if it is non-contextual via linear programming. Through Theorems 3 and 4 we have operational interpretations for the possible sequential HVMs.

VII. DISCUSSION AND CONCLUSION

A. Discussion

We now turn to the discussion of two references which have common objective with this paper, and show how they are related.

must exist $o' \in O_{\bar{S}}$ which is simply defined as $o' = o$

Relation to Ref [18]: The authors of Reference [18, 53] were looking at sequential measurements and the relation to non-contextuality with the assumption of both OD and non-invasiveness. Since this is related to our current work, we investigate and clarify the relation to this paper. Non-invasiveness means that the state is not disturbed by compatible measurements, which is a particular case of no-disturbance, where the transfer function is simply a Dirac function $\Gamma(\lambda'|\lambda, k) = \delta_{\lambda', \lambda}$. In [18], the authors propose multiple approaches to relax some notion of compatibility. We propose in the following to use only the weaker assumption of no-disturbance instead of the notion of compatibility they introduced for simplifying the notation. They relax the assumption of non-invasiveness by introducing the probability to flip, which they write as:

$$\begin{aligned} p^{\text{flip}}[AB] &= p[(b_1 = 1 | B_1) \text{ and } (b_2 = 0 | A_1 B_2)] \\ &\quad + p[(b_1 = 0 | B_1) \text{ and } (b_2 = 1 | A_1 B_2)]. \end{aligned}$$

Where the instrument B has outcome $\{0, 1\}$. We have interpreted the “and” as “at the same time”, which is why they say it is not a measurable quantity. Adapted to our notation this gives:

$$\begin{aligned} p^{\text{flip}}[AB] &= \sum_{\lambda_0, \lambda_1, k, k'} \mu(\lambda_0) \xi_B(k|\lambda_0) \\ &\quad \times \xi_A(k'|\lambda_0) \Gamma_A^\varepsilon(\lambda_1|\lambda_0, k') \\ &\quad \times \xi_B(1-k|\lambda_1) \end{aligned} \quad (\text{VII.1})$$

where $A \vdash B$. If the assumption of ND holds, it seems OD is never relaxed, then $p^{\text{flip}}[AB]$ would necessarily be 0. Because OD remains, the relaxation necessarily comes from the form of the transfer function Γ , thus one can introduce a noisy version:

$$\begin{aligned} \Gamma_A^\varepsilon(\lambda_1|\lambda_0, k') &= (1 - \varepsilon) \Gamma_A(\lambda_1|\lambda_0, k') \\ &\quad + \varepsilon \Gamma_A^{\text{noisy}}(\lambda_1|\lambda_0, k') \end{aligned} \quad (\text{VII.2})$$

where $\Gamma_A^{\text{noisy}}(\lambda_1|\lambda_0, k')$ is a function that chooses a λ_1 such that any non-disturbing measurement outcome is flipped. The existence of such function is not assured, however it follows the spirit of [18], and the notion of flipping. With this we have $p^{\text{flip}}[AB] = \varepsilon$.

However, as they argue, this might not be directly experimentally testable since both $\xi_B(k|\lambda_0)$ and $\xi_A(k'|\lambda_0)$ depend on λ_0 in Equation (VII.1) which is not possible because one has to chose which measurement comes first. We can

still follow through the paper with our approach, by introducing a new quantity $p^{\text{err}}[BAB]$:

$$p^{\text{err}}[BAB] = \sum_{\substack{\lambda_0, \lambda_1, \lambda_2 \\ k, k' \in \mathbb{B}}} \mu(\lambda_0) \xi_B(k|\lambda_0) \\ \times \Gamma_B(\lambda_1|\lambda_0, k) \xi_A(k'|\lambda_1) \text{(VII.3)} \\ \times \Gamma_A(\lambda_2|\lambda_1, k') \xi_B(1-k|\lambda_2)$$

where again $p^{\text{err}}[BAB] = 0$ as long as OD and ND are respected. The following inequality holds:

$$p^{\text{flip}}[AB] \leq p^{\text{err}}[BAB], \quad \text{(VII.4)}$$

as we have:

$$p^{\text{err}}[BAB] - p^{\text{flip}}[AB] = (1 - 2\varepsilon)\varepsilon \quad \text{(VII.5)}$$

and by definition $\varepsilon \leq 0.5$, otherwise it just amounts to flipping the outcomes of the measurements and take $\varepsilon' = 1 - \varepsilon$. As a conclusion, we acknowledge that there are more approaches used in [18] to account for errors. However, we have decided to restrict ourselves to the first one to give an example of how one can retrieve existing results with the framework we have developed in this paper.

Relation to [19] In a recent work Abramsky et al. [19] developed a framework for causal scenarios, with sequential measurements closely related to this work. The major differences are twofold, first we emphasize on hidden variable models, with state transformations and operational constraints which we develop. Second, we allow for repeated instruments, i.e. a sequence can have the same instrument multiple time, at the cost of getting away of the sheaf mathematical framework. Furthermore, in this paper, we investigate in more depth the notion of no-disturbance, and which constraints should one put on the HVMs in order to witness non-contextuality.

Yet we still believe that both works are likely complimentary, and it is left as an open avenue whether they can be unified in a more general framework.

B. Conclusion

To summarize, we have investigated sequential scenarios and their relation to non-contextuality. We first defined sequential scenarios and empirical behaviours, and subsequently we have looked at the definition of a HVM for sequential scenarios. We proposed some restrictions on this HVM

that led to an equivalence with the notion of non-contextuality. Furthermore, we also proved that the non-contextual set is a polytope for any sequential scenarios, with a well-defined notion of contextual fraction. Finally, we propose a way to map any measurement scenario to a sequential scenario.

Together this represents a powerful framework for considering non-classical behaviour in sequential scenarios. On the one hand we have a clear formal definition of the non-contextual polytope, which allows many tools to be applied, such as linear programming to certify the violation of an inequality given some experimental data [47]. On the other hand, in Theorems 3 and 4 we have clear interpretation that violation would imply the statistics cannot be explained by HVMs that are both non-disturbing and outcome deterministic for Theorem 3, or HVMs that are both non-disturbing and outcome independent for Theorem 4.

We leave open many avenues for future research in this work, amongst which, the relaxation determinism and no-disturbance, in the same spirit as [48]. We believe it is an interesting path since, as clearly pointed out in [18], the need for noise robust inequalities in sequential measurements is required to pursue realistic experiments. The hidden variable theory proposed in this work is a good starting point for such relaxations.

Another line of research follows along the line of operational contextuality [2], where one could see sequential measurements as a causal network [55], and see how that relates to operational contextuality.

Finally, it is not clear to the authors, how the mapping proposed in Section VI works for the quantum case, i.e., is any quantum behaviour in the sequential case mapped to a quantum behaviour in the Bell scenario? This goes along the lines of compiled XOR games [21, 22], but here the sequential scenario is defined differently, in particular related to what is called the ‘‘artificial constraints’’. We strongly believe that the reverse mapping does not exist since there is a variety of instruments that are acting non-trivially on the state, however this is still to be proven, and finding what exact subset it matches is of interest on its own.

VIII. ACKNOWLEDGEMENTS

We acknowledge the help of Matilde Baroni, Davide Rolino and Marco Túlio Quintino who indirectly improved many parts of this paper through informal discussions. We would also like to thank Gaël Massé who took the time to proof-read the paper and helped enhance the writing style. We acknowledge financial support from the French national quantum initiative managed by Agence Nationale de la Recherche in the framework of France 2030 through project EPIQ, with the reference ANR-22-PETQ-0007.

-
- [1] S. Kochen and E. P. Specker, The Problem of Hidden Variables in Quantum Mechanics, *Journal of Mathematics and Mechanics* **17**, 59 (1967), [24902153](#).
- [2] R. W. Spekkens, Contextuality for preparations, transformations, and unsharp measurements, *Physical Review A* **71**, 052108 (2005), [arxiv:quant-ph/0406166](#).
- [3] S. Abramsky and A. Brandenburger, The sheaf-theoretic structure of non-locality and contextuality, *New Journal of Physics* **13**, 113036 (2011).
- [4] E. N. Dzhafarov, J. V. Kujala, and V. H. Cervantes, *Contextuality-by-Default: A Brief Overview of Ideas, Concepts, and Terminology* (2015).
- [5] A. Acín, T. Fritz, A. Leverrier, and A. B. Sainz, A Combinatorial Approach to Nonlocality and Contextuality, *Communications in Mathematical Physics* **334**, 533 (2015).
- [6] A. Cabello, S. Severini, and A. Winter, Graph-Theoretic Approach to Quantum Correlations, *Physical Review Letters* **112**, 040401 (2014), [arxiv:1401.7081](#).
- [7] C. Budroni, A. Cabello, O. Gühne, M. Kleinmann, and J.-Å. Larsson, *Quantum Contextuality* (2022), [arxiv:2102.13036](#) [[math-ph](#), [physics:physics](#), [physics:quant-ph](#)].
- [8] R. Lapkiewicz, P. Li, C. Schaeff, N. K. Langford, S. Ramelow, M. Wieśniak, and A. Zeilinger, Experimental non-classicality of an indivisible quantum system, *Nature* **474**, 490 (2011).
- [9] J. Ahrens, E. Amselem, A. Cabello, and M. Bourennane, Two Fundamental Experimental Tests of Nonclassicality with Qutrits, *Scientific Reports* **3**, 2170 (2013).
- [10] X.-X. Chen, J. Li, M. Aili, Z. Meng, Y.-J. Wang, and A.-N. Zhang, Experimental demonstration of quantum contextuality with nine observables on a single photonic qutrit, *Physical Review A* **107**, 042431 (2023).
- [11] X. Zhang, M. Um, J. Zhang, S. An, Y. Wang, D.-l. Deng, C. Shen, L.-M. Duan, and K. Kim, State-Independent Experimental Test of Quantum Contextuality with a Single Trapped Ion, *Physical Review Letters* **110**, 070401 (2013).
- [12] R. Raussendorf, Contextuality in measurement-based quantum computation, *Physical Review A—Atomic, Molecular, and Optical Physics* **88**, 022322 (2013).
- [13] M. Howard, J. Wallman, V. Veitch, and J. Emerson, Contextuality supplies the ‘magic’ for quantum computation, *Nature* **510**, 351 (2014).
- [14] M. Lostaglio, Certifying quantum signatures in thermodynamics and metrology via contextuality of quantum linear response, *Phys. Rev. Lett.* **125**, 230603 (2020).
- [15] S. Gupta, D. Saha, Z.-P. Xu, A. Cabello, and A. S. Majumdar, Quantum contextuality provides communication complexity advantage, *Phys. Rev. Lett.* **130**, 080802 (2023).
- [16] K. Horodecki, M. Horodecki, P. Horodecki, R. Horodecki, M. Pawłowski, and M. Bourennane, Contextuality offers device-independent security, *arXiv preprint arXiv:1006.0468* (2010).
- [17] R. W. Spekkens, D. H. Buzacott, A. J. Keehn, B. Toner, and G. J. Pryde, Preparation contextuality powers parity-oblivious multiplexing, *Physical review letters* **102**, 010401 (2009).
- [18] O. Gühne, M. Kleinmann, A. Cabello, J.-A. Larsson, G. Kirchmair, F. Zähringer, R. Gerritsma, and C. F. Roos, Compatibility and noncontextuality for sequential measurements, *Physical Review A* **81**, 022121 (2010), [arxiv:0912.4846](#).
- [19] S. Abramsky, R. S. Barbosa, and A. Searle, Combining contextuality and causality: A game semantics approach (2023), [2307.04786](#) [[quant-ph](#)].
- [20] V. J. Wright and M. Farkas, Invertible Map between Bell Nonlocal and Contextuality Scenarios, *Physical Review Letters* **131**, 220202 (2023).
- [21] L. Catani, R. Faleiro, P.-E. Emeriau, S. Mansfield, and A. Pappa, *Connecting XOR and XOR* games* (2024), [arXiv:2210.00397](#).
- [22] M. Baroni, Q.-H. Vu, B. Bourdoncle, E. Diamanti, D. Markham, and I. Šupić, *Quantum bounds for compiled XOR games and $\$d\$$ -outcome CHSH games* (2024), [arXiv:2403.05502](#) [[quant-ph](#)].
- [23] M. F. Pusey and M. S. Leifer, Logical pre- and post-selection paradoxes are proofs of contextuality, *Electronic Proceedings in Theoretical Computer Science* **195**, 295 (2015), [arXiv:1506.07850](#) [[quant-ph](#)].
- [24] M. S. Leifer and R. W. Spekkens, Logical Pre- and Post-Selection paradoxes, measurement-disturbance and contextuality, *International Journal of Theoretical Physics* **44**, 1977 (2005), [arXiv:quant-ph/0412179](#).
- [25] M. Plávala, General probabilistic theories: An introduction, *Physics Reports General Probabilistic Theories: An Introduction*, **1033**, 1 (2023).
- [26] T. Heinosaari, L. Leppäjärvi, and M. Plávala, No-free-information principle in general probabilistic theories, *Quantum* **3**, 157 (2019), [arXiv:1808.07376](#) [[quant-ph](#)].
- [27] G. M. D’Ariano, G. Chiribella, and P. Perinotti, *Quantum Theory from First Principles: An Informational Approach* (Cambridge University Press, Cambridge, 2017).
- [28] G. M. D’Ariano, P. Perinotti, and A. Tosini, Incompatibility of observables, channels and in-

- struments in information theories, *Journal of Physics A: Mathematical and Theoretical* **55**, 394006 (2022), [arXiv:2204.07956 \[quant-ph\]](#).
- [29] F. Buscemi, K. Kobayashi, S. Minagawa, P. Perinotti, and A. Tosini, Unifying different notions of quantum incompatibility into a strict hierarchy of resource theories of communication, *Quantum* **7**, 1035 (2023).
- [30] T. Heinosaari and M. M. Wolf, Nondisturbing quantum measurements, *Journal of Mathematical Physics* **51**, 092201 (2010).
- [31] A. Cabello, S. Severini, and A. Winter, (Non-)Contextuality of Physical Theories as an Axiom, [arXiv:1010.2163 \[quant-ph\]](#) (2010), [arXiv:1010.2163 \[quant-ph\]](#).
- [32] R. Ramanathan, A. Soeda, P. Kurzyński, and D. Kaszlikowski, Generalized Monogamy of Contextual Inequalities from the No-Disturbance Principle, *Physical Review Letters* **109**, 050404 (2012).
- [33] J. Kofler and Č. Brukner, Condition for macroscopic realism beyond the Leggett-Garg inequalities, *Physical Review A* **87**, 052115 (2013).
- [34] P. Busch, "No Information Without Disturbance": Quantum Limitations of Measurement (2009), [arXiv:0706.3526 \[quant-ph\]](#).
- [35] T. Heinosaari, T. Miyadera, and M. Ziman, An invitation to quantum incompatibility, *Journal of Physics A: Mathematical and Theoretical* **49**, 123001 (2016).
- [36] A. A. Klyachko, M. A. Can, S. Binicioğlu, and A. S. Shumovsky, Simple Test for Hidden Variables in Spin-1 Systems, *Physical Review Letters* **101**, 020403 (2008).
- [37] H. Dourdent, *Contextuality, Witness of Quantum Weirdness* (2018), [1801.09768 \[quant-ph\]](#).
- [38] Y. Guryanova, R. Silva, A. J. Short, P. Skrzypczyk, N. Brunner, and S. Popescu, Exploring the limits of no backwards in time signalling, *Quantum* **3**, 211 (2019).
- [39] E. B. Davies, *Quantum Theory of Open Systems*, academic press ed. (Academic Press, 1976).
- [40] J. S. Bell, On the Einstein Podolsky Rosen paradox, *Physics Physique Fizika* **1**, 195 (1964).
- [41] T. Norsen, John S. Bell's concept of local causality, *American Journal of Physics* **79**, 1261 (2011).
- [42] A. Fine, Hidden Variables, Joint Probability, and the Bell Inequalities, *Physical Review Letters* **48**, 291 (1982).
- [43] M. D. Mazurek, M. F. Pusey, R. Kunjwal, K. J. Resch, and R. W. Spekkens, An experimental test of noncontextuality without unphysical idealizations, *Nature Communications* **7**, ncomms11780 (2016).
- [44] M. F. Pusey, L. del Rio, and B. Meyer, *Contextuality without access to a tomographically complete set* (2019), [arXiv:1904.08699 \[quant-ph\]](#).
- [45] V. Gitton and M. P. Woods, Solvable Criterion for the Contextuality of any Prepare-and-Measure Scenario, *Quantum* **6**, 732 (2022).
- [46] D. Schmid, J. H. Selby, E. Wolfe, R. Kunjwal, and R. W. Spekkens, Characterization of Non-contextuality in the Framework of Generalized Probabilistic Theories, *PRX Quantum* **2**, 010331 (2021).
- [47] S. Abramsky, R. S. Barbosa, and S. Mansfield, The contextual fraction as a measure of contextuality, *Physical Review Letters* **119**, 050504 (2017), [arXiv:1705.07918](#).
- [48] K. Vallée, P.-E. Emeriau, B. Bourdoncle, A. Sobhi, S. Mansfield, and D. Markham, Corrected Bell and non-contextuality inequalities for realistic experiments, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **382**, 20230011 (2024).
- [49] A. Peres, Incompatible results of quantum measurements, *Physics Letters A* **151**, 107 (1990).
- [50] N. D. Mermin, Hidden variables and the two theorems of John Bell, *Reviews of Modern Physics* **65**, 803 (1993).
- [51] P. Busch and P. Lahti, Lüders Rule, in *Compendium of Quantum Physics*, edited by D. Greenberger, K. Hentschel, and F. Weinert (Springer, Berlin, Heidelberg, 2009) pp. 356–358.
- [52] A. Cabello, P. Badziąg, M. Terra Cunha, and M. Bourennane, Simple Hardy-Like Proof of Quantum Contextuality, *Physical Review Letters* **111**, 180404 (2013).
- [53] G. Kirchmair, F. Zähringer, R. Gerritsma, M. Kleinmann, O. Gühne, A. Cabello, R. Blatt, and C. F. Roos, State-independent experimental test of quantum contextuality, *Nature* **460**, 494 (2009), [arXiv:0904.1655](#).
- [54] M. Araújo, M. T. Quintino, C. Budroni, M. T. Cunha, and A. Cabello, All noncontextuality inequalities for the n -cycle scenario, *Physical Review A* **88**, 022118 (2013).
- [55] R. Chaves, G. Moreno, E. Polino, D. Poderini, I. Agresti, A. Suprano, M. R. Barros, G. Carvacho, E. Wolfe, A. Canabarro, R. W. Spekkens, and F. Sciarrino, Causal Networks and Freedom of Choice in Bell's Theorem, *PRX Quantum* **2**, 040323 (2021).

Appendix A: Proofs

1. Proof of Theorem 2

Proof. Let an arbitrary hidden behaviour h from a ND HVM $\langle \Lambda, \mu, (h^\lambda)_{\lambda \in \Lambda} \rangle$ on a NDS scenario $\langle X, \mathcal{S}, \mathcal{O} \rangle_{\text{seq}}$, any two contexts $S, S' \in \mathcal{S}$, and define $N = |S|$ and $N' = |S'|$ such that:

$$S = \{A_k\}_{1 \leq k \leq N, A_k \in \mathcal{S}} \quad (\text{A.1})$$

$$S' = \{A_\ell\}_{1 \leq \ell \leq N', A_\ell \in \mathcal{S}'}, \quad (\text{A.2})$$

Let us take the case where:

$$\exists \tilde{k}, \tilde{\ell}: A_{\tilde{k}} \in S: A_{\tilde{\ell}} \in S': A_{\tilde{k}} = A_{\tilde{\ell}},$$

We now check that the compatibility of marginal is respected:

$$p(a_{\tilde{k}} | A_{\tilde{k}}, S, \lambda_0) = \sum_{\substack{A_k \in S \\ a_k \neq a_{\tilde{k}}}} p(a_1, \dots, a_N | A_1, \dots, A_N) \quad (\text{A.3})$$

$$= \sum_{\substack{\lambda_1, \dots, \lambda_{N+1} \\ a_k \neq a_{\tilde{k}}}} \prod_{k=1}^{N+1} \xi_{A_k}(a_k | \lambda_{k-1}) \Gamma_{A_k}(\lambda_k | \lambda_{k-1}, a_k) \quad (\text{A.4})$$

$$= \sum_{\substack{\lambda_1, \dots, \lambda_{N+1} \\ a_k \neq a_{\tilde{k}}}} \left[\prod_{k=1}^{\tilde{k}-1} \xi_{A_k}(a_k | \lambda_{k-1}) \Gamma_{A_k}(\lambda_k | \lambda_{k-1}, a_k) \right] \\ \times \xi_{A_{\tilde{k}}}(a_{\tilde{k}} | \lambda_{\tilde{k}-1}) \Gamma_{A_{\tilde{k}}}(\lambda_{\tilde{k}} | \lambda_{\tilde{k}-1}, a_{\tilde{k}}) \\ \times \left[\prod_{k=\tilde{k}+1}^{N+1} \xi_{A_k}(a_k | \lambda_{k-1}) \Gamma_{A_k}(\lambda_k | \lambda_{k-1}, a_k) \right] \quad (\text{A.5})$$

$$= \sum_{\substack{\lambda_1, \dots, \lambda_{\tilde{k}-1} \\ a_k \neq a_{\tilde{k}}}} \left[\prod_{k=1}^{\tilde{k}-1} \xi_{A_k}(a_k | \lambda_{k-1}) \Gamma_{A_k}(\lambda_k | \lambda_{k-1}, a_k) \right] \\ \times \xi_{A_{\tilde{k}}}(a_{\tilde{k}} | \lambda_{\tilde{k}-1}) \quad (\text{A.6})$$

$$= \sum_{\substack{\lambda_1, \dots, \lambda_{\tilde{k}-1} \\ a_k \neq a_{\tilde{k}}}} \left[\prod_{k=1}^{\tilde{k}-2} \xi_{A_k}(a_k | \lambda_{k-1}) \Gamma_{A_k}(\lambda_k | \lambda_{k-1}, a_k) \right] \\ \times \xi_{A_{\tilde{k}-1}}(a_{\tilde{k}-1} | \lambda_{\tilde{k}-2}) \Gamma_{A_{\tilde{k}-1}}(\lambda_{\tilde{k}-1} | \lambda_{\tilde{k}-2}, a_{\tilde{k}-1}) \xi_{A_{\tilde{k}}}(a_{\tilde{k}} | \lambda_{\tilde{k}-1}) \quad (\text{A.7})$$

$$= \sum_{\substack{\lambda_1, \dots, \lambda_{\tilde{k}-2} \\ a_k \neq a_{\tilde{k}}}} \left[\prod_{k=1}^{\tilde{k}-2} \xi_{A_k}(a_k | \lambda_{k-1}) \Gamma_{A_k}(\lambda_k | \lambda_{k-1}, a_k) \right] \\ \times \xi_{A_{\tilde{k}}}(a_{\tilde{k}} | \lambda_{\tilde{k}-2}) \quad (\text{A.8})$$

$$= \xi_{A_{\tilde{k}}}(a_{\tilde{k}} | \lambda_0) \quad (\text{A.9})$$

where A_k is the k^{th} measurement of context S . In Equation (A.4), we used NBTS condition. Then we split the instruments that come before and after $A_{\bar{k}}$ in Equation (A.5), and because of the summation we can directly eliminate the second product in Equation (A.6), finally we just iteratively apply ND in Equation (A.7) and (A.8). We can apply iteratively the ND relation because by definition of a ND HVM we have:

$$\forall j < m: A_j \vdash A_m \quad (\text{A.10})$$

We can also apply this procedure for the context S' , and it is quite clear that we obtain the same result, proving:

$$p(a_{\bar{k}}|A_{\bar{k}}, S, \lambda_0) = \xi_{A_{\bar{k}}}(a_{\bar{k}}|\lambda_0) \quad (\text{A.11})$$

$$= \xi_{A_{\bar{\ell}}}(a_{\bar{\ell}}|\lambda_0) \quad (\text{A.12})$$

$$= p(a_{\bar{\ell}}|A_{\bar{\ell}}, S', \lambda_0) \quad (\text{A.13})$$

hence we do have compatibility of marginal. \square

2. Proof of Theorem 3

In order to prove Theorem 3, we state the following useful lemma:

Lemma 2. *Let two deterministic instruments A and A' performed in sequence such that $A \vdash A'$. Then the following holds:*

$$\begin{aligned} \forall \lambda \in \Lambda: \forall \lambda' \in \text{supp} \{ \Gamma_A(\cdot|\lambda, a) \}: \forall a' \in O_{A'}: \\ \xi_{A'}(a'|\lambda') = \xi_{A'}(a'|\lambda) \end{aligned} \quad (\text{A.14})$$

where $\text{supp} \{ \Gamma_A(\cdot|\lambda, a) \}$ stands for the support of the transfer function Γ_A , i.e., when Γ_A is non-zero.

Proof. Let us denote by $a_0^{\lambda_0}$ the outcome such that $\xi_{A_0}(a_0|\lambda_0) = 1$. The existence and uniqueness of such an outcome is given by the OD assumption. By using Equation (III.2) on no-disturbance we obtain:

$$\forall \lambda_0: \forall a_1: \sum_{\lambda_1} \Gamma_{A_0}(\lambda_1|\lambda_0, a_0^{\lambda_0}) \xi_{A_1}(a_1|\lambda_1) = \xi_{A_1}(a_1|\lambda_0) \quad (\text{A.15})$$

We now proceed by enumeration, since $\xi_{A_1}(a_1|\lambda_0) \in \{0, 1\}$. First we assume $\xi_{A_1}(a_1|\lambda_0) = 0$, then necessarily

$$\forall \lambda_1 \in \text{supp}(\Gamma_{A_0}): \xi_{A_1}(a_1|\lambda_1) = 0 \quad (\text{A.16})$$

otherwise we would have

$$\Gamma_{A_0}(\lambda_1|\lambda_0, a_0^{\lambda_0}) \xi_{A_1}(a_1|\lambda_1) > 0 \neq \xi_{A_1}(a_1|\lambda_0). \quad (\text{A.17})$$

We can extend similarly for $\xi_{A_1}(a_1|\lambda_0) = 1$:

$$\forall \lambda_1 \in \text{supp}(\Gamma_{A_0}): \xi_{A_1}(a_1|\lambda_1) = 1 \quad (\text{A.18})$$

Otherwise we have

$$\sum_{\lambda_1} \Gamma_{A_0}(\lambda_1|\lambda_0, a_0^{\lambda_0}) \xi_{A_1}(a_1|\lambda_1) < 1 \quad (\text{A.19})$$

$$= \xi_{A_1}(a_1|\lambda_0) \quad (\text{A.20})$$

\square

We now turn to the proof of the main theorem:

Proof. Let an empirical model e on a sequential scenario $\langle X, \mathcal{S}, O \rangle_{\text{seq}}$ and let any context $S \in \mathcal{S}$ with $N = |S|$. In full generality, we denote the instruments in the context S by A_i where i is the order of the instrument, and each instrument is in X . Suppose now that we have an OD ND HVM $\langle \Lambda, \mu, (h^\lambda)_{\lambda \in \Lambda} \rangle$. Then we write $o \in O_S$ as $o = (a_i)_{i \leq N}$, and we have:

$$e_S(o) = \sum_{\lambda_0, \dots, \lambda_{N+1}} \mu(\lambda_0) \prod_{j=1}^N \xi_{A_j}(a_j | \lambda_{j-1}) \Gamma_{A_j}(\lambda_j | \lambda_{j-1}, a_j) \quad (\text{A.21})$$

where $\forall j = 1, \dots, N$ we have a_j to be the outcome of measurement A_j . With some rearranging, we obtain:

$$e_S(o) = \sum_{\lambda_0, \dots, \lambda_N} \mu(\lambda_0) \xi_{A_1}(a_1 | \lambda_0) \prod_{j=2}^N \Gamma_{A_{j-1}}(\lambda_{j-1} | \lambda_{j-2}, a_{j-1}) \xi_{A_j}(a_j | \lambda_{j-1}) \quad (\text{A.22})$$

We now use lemma 2, because we have for all j that $A^{j-1} \vdash A^j$ we can rewrite the above equation as:

$$\begin{aligned} e_S(o) &= \sum_{\lambda_0, \dots, \lambda_N} \mu(\lambda_0) \xi_{A_1}(a_1 | \lambda_0) \xi_{A_2}(a_2 | \lambda_0) \\ &\quad \times \left[\prod_{j=2}^{N-1} \Gamma_{A_{j-1}}(\lambda_{j-1} | \lambda_{j-2}, a_{j-1}) \xi_{A_{j+1}}(a_{j+1} | \lambda_j) \right] \\ &\quad \times \Gamma_{A_{N-1}}(\lambda_{N-1} | \lambda_{N-2}, a_{N-1}) \end{aligned} \quad (\text{A.23})$$

By acting iteratively we end up with:

$$e_S(o) = \sum_{\lambda_0, \dots, \lambda_N} \mu(\lambda_0) \left[\prod_{j=1}^N \xi_{A_j}(a_j | \lambda_0) \right] \left[\prod_{j=1}^{N-1} \Gamma_{A_j}(\lambda_j | \lambda_{j-1}, a_j) \right] \quad (\text{A.24})$$

Where the transfer functions Γ_{A_j} can be eliminated by the sum over the right λ_j :

$$e_S(o) = \sum_{\lambda_0} \mu(\lambda_0) \left[\prod_{j=1}^N \xi_{A_j}(a_j | \lambda_0) \right] \quad (\text{A.25})$$

We now show that there exist a global probability distribution on outcomes $d: O_X \rightarrow [0, 1]$ which respects the definition of non-contextuality given in Definition 13. First, if o is not consistent, it means that there exists two instruments A_i, A_j which are the same $A_i = A_j = A$ but which have different outcomes $a_i \neq a_j$. This implies $\xi_A(a_i | \lambda_0) \xi_A(a_j | \lambda_0) = 0$ since we have deterministic response functions one of them must be 0. In terms of the value of $e_S(o)$ it implies:

$$\begin{aligned} e_S(o) &\propto \xi_{A_i}(a_i | \lambda_0) \xi_{A_j}(a_j | \lambda_0) \\ &\propto 0. \end{aligned} \quad (\text{A.26})$$

Thus, necessarily $e_S(o) = 0$ for inconsistent outcomes. For consistent outcomes we can rigorously follow the proof in [3, Theorem 8.1]. Let us define the global distribution on an outcome $o \in O_X$, which we write as $o = (a)_{A \in X}$:

$$d(o) := \sum_{\lambda_0 \in \Lambda} \mu(\lambda_0) \prod_{A \in X} \xi_A(a | \lambda_0) \quad (\text{A.27})$$

We naturally have that $d(o) \geq 0$, and it normalizes to one:

$$\begin{aligned}
\sum_{o \in O_X} d(o) &= \sum_{o \in O_X} \sum_{\lambda_0 \in \Lambda} \mu(\lambda_0) \prod_{A \in X} \xi_A(a|\lambda_0) \\
&= \sum_{\lambda_0 \in \Lambda} \mu(\lambda_0) \prod_{A \in X} \sum_a \xi_A(a|\lambda_0) \\
&= \sum_{\lambda_0 \in \Lambda} \mu(\lambda_0) \\
&= 1
\end{aligned} \tag{A.28}$$

We finally need to check that it marginalizes to the good value for a given context $S \in \mathcal{S}$:

$$e_S(o) = \sum_{\lambda_0} \mu(\lambda_0) \left[\prod_{j=1}^N \xi_{A_j}(a_j|\lambda_0) \right] \tag{A.29}$$

$$= \sum_{\lambda_0} \mu(\lambda_0) \prod_{A \in \bar{S}} \xi_A(a|\lambda_0) \tag{A.30}$$

$$= d|_{\bar{S}}(o_{\bar{S}}) \tag{A.31}$$

where we went from Equation (A.29) to Equation (A.30) by using the fact that $\xi_A^2(a|\lambda_0) = \xi_A(a|\lambda_0)$ for all $A \in X$, thus ending with non-repeating measurements only.

The proof for the converse statement, i.e., the existence of a hidden variable model given a global probability distribution, follows naturally from [3, Prop. 3.1], where now for all $S \in \mathcal{S}$ we set

$$e_S(o) = \begin{cases} d|_{\bar{S}}(o_{\bar{S}}), & \text{if } o \text{ is consistent} \\ 0, & \text{otherwise} \end{cases} \tag{A.32}$$

Hence, $e_S(o)$ is explained by a factorizable hidden variable model. Then any transfer function that induce no-disturbance, such as Dirac functions. For repeating measurements, one can use the deterministic property that $\xi_A(a|\lambda) = [\xi_A(a|\lambda)]^n$ for any $n \in \mathbb{N}^+$. \square

3. Proof of Theorem 4

We start by showing that the notion of ND simplifies with the assumption of OI:

Lemma 3 (ND with the assumption of OI). *Let and OI ND HVM $\langle \Lambda, \mu, (h^\lambda)_{\lambda \in \Lambda} \rangle$, then for any $A_1 \vdash A_2$ the following holds:*

$$\begin{aligned}
&\forall \lambda_0 : \forall a_1 : \\
&\sum_{\lambda_1} \Gamma_{A_1}(\lambda_1|\lambda_0) \xi_{A_2}(a_2|\lambda_1) = \xi_{A_2}(a_2|\lambda_0)
\end{aligned} \tag{A.33}$$

Proof. This comes quite straightforwardly from Definition 9, where the first term $\xi_{A_i}(a_i|\lambda_0)$ is ruled out by summing over a_1 since the transfer function is no more outcome dependent. \square

We now turn to the proof of Theorem 4.

Proof. Let an OI ND HVM $\langle \Lambda, \mu, (h^\lambda)_{\lambda \in \Lambda} \rangle$ on a sequential scenario $\langle X, \mathcal{S}, O \rangle_{\text{seq}}$, then we will show that this takes the form of a factorizable HVM. Let $S \in \mathcal{S}$ and two measurements $A_1, A_2 \in S$, by definition, we have $A_1 \vdash A_2$ and:

$$\begin{aligned}
p(a_1, a_2|A_1, A_2) &= \sum_{\lambda_0, \lambda_1} \mu(\lambda_0) \xi_{A_1}(a_1|\lambda_0) \Gamma_{A_1}(\lambda_1|\lambda_0, a_1) \xi_{A_2}(a_2|\lambda_1) \\
&= \sum_{\lambda_0} \mu(\lambda_0) \xi_{A_1}(a_1|\lambda_0) \xi_{A_2}(a_2|\lambda_0)
\end{aligned} \tag{A.34}$$

Where we went from the first line to the second by using Lemma 3. Because this is a factorizable HVM, we can follow the same reasoning as applied in the proof of Theorem 3, in Appendix A 2. \square