Beyond the Cabello-Severini-Winter framework:
Making sense of contextuality without sharpness of measurements

Ravi Kunjwal
Perimeter Institute for Theoretical Physics,
31 Caroline Street North, Waterloo,
Ontario, Canada, N2L 2Y5.
(Dated: February 22, 2018)

By generalizing the Cabello-Severini-Winter (CSW) framework, we build a bridge from this graph-theoretic approach for Kochen-Specker (KS) contextuality to a hypergraph-theoretic approach for Spekkens’ contextuality, as applied to Kochen-Specker type scenarios. The CSW framework focuses on the correlations obtained from compatible sets of measurements carried out on a system prepared according to a fixed preparation procedure (as in Bell-KS type experiments). On the other hand, our generalized framework describes an experiment that requires, besides the correlations considered by CSW, the correlations between measurement events and corresponding preparation procedures that seek to make these measurement events highly predictable. This latter feature of the experiment allows us to obtain noise-robust noncontextuality inequalities by applying the assumption of noncontextuality to both preparations and measurements, without requiring the assumption of outcome determinism. Indeed, we treat all measurements on an equal footing: no notion of “sharpness” is presumed for them, hence no putative justification of outcome determinism from sharpness is sought. As a result, unlike the CSW framework, we do not require Specker’s principle (also known as the exclusivity principle) — that pairwise exclusive measurement events must all be mutually exclusive — as a fundamental constraint on (sharp) measurement events in any operational theory describing the experiment. All this allows us, for the case of quantum theory, to deal with nonprojective (or unsharp) measurements without running into a crucial pathology of approaches based on Kochen-Specker contextuality, such as that of CSW: that trivial POVMs (positive operator-valued measures) can lead to maximal violations of KS-noncontextuality inequalities, thus being seemingly “maximally nonclassical” by the lights of KS-contextuality. In our approach, trivial POVMs never lead to a violation of our noncontextuality inequalities, thus never manifesting any contextuality, i.e., nonclassicality by the lights of Spekkens’ contextuality. This feature arises as a natural consequence of the framework of generalized noncontextuality rather than being put in “by hand”. Our noncontextuality inequalities are therefore robust to the presence of noise in the experimental procedures, whether they are measurements or preparations, and are applicable to operational theories that need not be quantum.

PACS numbers: 03.65.Ta, 03.65.Ud

CONTENTS

I. Introduction 2

II. Spekkens framework 4

III. Hypergraph approach to Kochen-Specker scenarios in the Spekkens framework 6

A. Measurements

1. Classification of probabilistic models 7

2. Structural Specker’s principle vs. Statistical Specker’s principle 8

3. What does it mean for an operational theory to satisfy structural/statistical Specker’s principle? 9

4. Remark on the classification of probabilistic models: why we haven’t defined “quantum models” as those obtained from projective measurements 10

B. Sources 11

IV. A key hypergraph invariant: the weighted max-predictability 14

V. Noise-robust noncontextuality inequalities 14

A. Example: KCBS scenario 16

VI. Discussion 18

A. Measurement-measurement correlations vs. source-measurement correlations 18

B. Can our noise-robust noncontextuality inequalities be saturated by a noncontextual ontological model? 18

C. Can trivial POVMs ever violate these noncontextuality inequalities? 19

D. Open questions 20

VII. Conclusions 21

Acknowledgments 21

*rkunjwal@perimeterinstitute.ca
I. INTRODUCTION

Much work has been devoted, recently [1–6], to obtaining constraints on operational statistics that follow from the assumption of noncontextuality within the framework proposed by Spekkens [7]. This generalized framework abandons the assumption of outcome determinism that is intrinsic to the Kochen-Specker (KS) framework [8], applies to arbitrary operational theories, and extends the notion of noncontextuality to arbitrary experimental procedures – preparations, transformations, and measurements – rather than measurements alone.

On the other hand, work along the lines of the traditional KS framework culminated in two recent approaches: the graph-theoretic framework of Cabello, Severini, and Winter (CSW) [9, 10] where a general approach to obtaining graph-theoretic bounds on linear Bell-KS functionals was proposed, and the related hypergraph framework of Acín, Fritz, Leverrier, and Sainz (AFLS) [11], where an approach to characterizing sets of correlations was proposed. The CSW framework relates well-known graph invariants to upper bounds of Bell-KS inequalities, upper bounds on maximum quantum violations of these inequalities, and upper bounds on them in general probabilistic theories [12] – denoted $E_1$ – which satisfy the “exclusivity principle” [10]. Complementary to this, the AFLS framework uses graph invariants in the service of deciding whether a given assignment of probabilities to measurement outcomes in a KS-contextuality experiment belongs to a particular set of correlations; they showed that membership in the quantum set of correlations (defined only for projective measurements in quantum theory) cannot be witnessed by a graph invariant [11]. Another recent approach [13] employs sheaf-theoretic ideas to formulate KS-contextuality.

In this paper we build a bridge from the CSW approach, where “classical” (i.e., KS-noncontextual) correlations are bounded by Bell-KS inequalities, to noise-robust noncontextuality inequalities in the Spekkens framework [7]. Unlike the criteria for KS-contextuality in the CSW framework, the operational criteria for contextuality à la Spekkens are robust to noise and therefore applicable to arbitrary positive operator-valued measures (POVMs) and mixed states in quantum theory.

Indeed, if one allows for POVMs in the definition of quantum correlations (rather than just projective measurements), then the separation between quantum and $E_1$ correlations in the CSW framework breaks down. This is because any set of probabilities satisfying the “no-disturbance” or “no-signalling” condition (of which the $E_1$ correlations are a subset, in general) can be achieved by (trivial) POVMs by simply multiplying an identity operator with every probability in such an assignment of probabilities[1]. By the lights of KS-noncontextuality as one’s notion of classicality, then, trivial POVMs saturating the general probabilistic bound on the correlations are maximally nonclassical (i.e., maximally KS-contextual). However, we “know” intuitively that trivial POVMs are “classical”, even if KS-noncontextuality as a notion of classicality doesn’t quite capture that intuition. A simple operational sense in which trivial POVMs are “classical” is that they reveal nothing about the quantum state on which they are measured, being incapable of distinguishing any pair of states whatsoever. This simple sense in which they are “classical” is, however, not captured by KS-noncontextuality as one’s notion of classicality, since the experiment is restricted to considering correlations obtained from compatible sets of measurements implemented on the same preparation, and therefore no variation over preparations is taken into account in Bell-KS type inequalities. This makes such experiments incapable of witnessing the “triviality” of trivial POVMs, i.e., the fact that they correspond to a fixed probability distribution that doesn’t vary even as the choice of preparation is varied. Moreover, since all nonprojective measurements are excluded by fiat in traditional Kochen-Specker type approaches [10, 11] one loses out on the potential to explore the possibilities that nontrivial, yet nonprojective, measurements offer with respect to contextuality[3].

Note that whenever we refer to “Bell-KS” functionals or inequalities for Kochen-Specker type experiments, we are not thinking of experiments that are Bell experiments [18, 23], which have spacelike separation between multiple parties, each performing local measurements on a shared multipartite preparation. For the case of Bell experiments, trivial local POVMs assigned to each party in a Bell experiment do not lead to Bell violations for a simple reason: the trivial POVMs for each party are all compatible with each other, thereby admitting a joint probability distribution over their outcomes for each party; taking a product of these local joint probability distributions (one for each party) results in a joint distribution over all measurements of all parties, hence satisfying Bell

---

1. Trivial POVMs are, therefore, trivial resolutions of the identity, where every POVM element is proportional to identity, i.e., \( \{ a I \}_a \) such that \( a \in [0,1] \) and \( \sum_a a = 1 \).

2. Indeed, any trivial POVM can be realized in the following operational manner: take the quantum system prepared in some state, throw it in the garbage, and then sample from the classical probability distribution corresponding to the trivial POVM.

3. Partly to avoid the pathology with respect to trivial POVMs that we have just pointed out.

4. All trivial POVMs are nonprojective, but not all nonprojective POVMs are trivial. Indeed, see Refs. [13, 17] for examples of generalized contextuality [7] with nonprojective measurements, albeit assuming operational quantum theory.
inequalities. The fact that the POVMs are trivial ensures that the Bell inequalities are satisfied regardless of the choice of shared quantum state. On the other hand, forgetting the constraint of local POVMs, there always exist global trivial POVMs that can violate Bell inequalities: e.g., just take the Popescu-Rohrlich (PR) box distribution [24], and multiply an identity operator (on the joint Hilbert space of Alice and Bob) with each probability in the PR-box; this results in four trivial POVMs, defined over the joint Hilbert space, that together violate the CHSH inequality maximally. But, of course, this violation is uninteresting because it doesn’t obey the locality constraint on the measurements in a Bell experiment. This is mathematically reflected in the fact that the PR-box distribution cannot be written as a convex mixture of product distributions, one for each party, hence the corresponding trivial POVM cannot be understood in terms of trivial local POVMs. Hence, it is the locality of the trivial POVMs in a Bell experiment that prevents them from violating a Bell inequality. The fact that they are “trivial” in the sense of being unable to distinguish two quantum states plays a role in the sense that, regardless of the shared quantum state, these POVMs yield fixed distributions over the measurement outcomes, thus always allowing the construction of a fixed (that is, independent of the quantum state) global joint probability distribution over all measurements in a Bell scenario. Since there are no such locality constraints on the form of the POVM elements in a Kochen-Specker experiment, they can easily violate any KS-noncontextuality inequality, e.g., the two-party CHSH experiment considered as a Kochen-Specker experiment with four observables in a 4-cycle where adjacent pairs are jointly measurable allows for trivial POVMs (like the PR-box trivial POVM above) violating the CHSH-type Bell-KS inequality in this scenario maximally. By the lights of KS-noncontextuality, this violation would indicate the maximum possible KS-contextuality with respect to this CHSH-type inequality.

Hence, our criticism of KS-noncontextuality as a notion of classicality — in an experiment with no locality constraints on the measurements — does not extend to the case of Bell-locality (or local causality) as a notion of classicality in a Bell experiment, where the experiment must respect locality constraints on the measurements for a Bell inequality violation to be meaningful. It is the locality of the measurements implemented by the various parties in a Bell experiment that renders Bell-locality immune to the criticism we are directing at KS-noncontextuality in this paper. Indeed, any attempt at a unified approach to KS-contextuality and Bell-nonlocality in the traditional approach [10] suffers from the problem of not making this distinction (one of locality of POVMs) between the two kinds of experiments (Bell experiments vs. Kochen-Specker experiments) precise, choosing instead to dwell on their formal mathematical unification as an instance of the classical marginal problem [23]. The marginal problem formulation is perhaps most explicit in the case of marginal scenarios defined in Ref. [26] (see also Ref. [27]). This unification forces a certain dichotomy in these approaches: while in Bell scenarios, one need not restrict to any notion of a “sharp” measurement in the definition of probabilistic models (and thus claim “theory independence”), in Kochen-Specker scenarios, one must make some statement about the structure of the measurements (such as their presumed sharpness [28], or that their joint measurability [17] is restricted to commutativity [13]), rendering any putative “theory independence” claim unfounded.

Because of this pathology of POVMs with respect to KS-noncontextuality as a notion of classicality, all traditional treatments in the KS framework [10] restrict the set of quantum correlations to those which are achieved by projective measurements (rather than POVMs, generally) on a quantum state. With recent work on a sensible notion of “sharp” measurement in a general probabilistic theory [31], the current attitude of proponents (see, e.g., [28]) of the traditional KS framework (defending KS-noncontextuality as a sensible operational notion of classicality) is to restrict attention to sharp measurements in both quantum theory and general probabilistic theories. However, another logical possibility is available and, indeed, operationally better justified than KS-noncontextuality [33]: that one must revise one’s notion of classicality as KS-noncontextuality to a notion of classicality that allows for arbitrary quantum measurements and witnesses the fact that trivial POVMs are indeed classical according to this revised notion, even in experiments where — unlike a Bell experiment — there is no constraint of locality on the measurements. At the same time, such a revised notion should be capable of recovering the traditional notion of KS-noncontextuality as classicality in the case of projective measurements in quantum theory. Fortunately, we already have such a notion of classicality: namely, universal noncontextuality, as defined in the Spekkens framework [7]. In

---

6 See Ref. [30] for how this lack of locality of measurements in a Kochen-Specker type experiment translates, at the ontological level, to the unreasonableness of assuming factorizability in the ontological model; this factorizability (or the stronger condition of outcome determinism) is invoked to justify the resulting derivation of Bell-KS inequalities as constraints from a classical marginal problem.

7 This is in contrast to what is usually done traditionally: that one insists on KS-noncontextuality as one’s notion of classicality [10] and, for this notion to make sense, one restricts the scope of allowed measurements to just the projective measurements so that commutativity is equivalent to joint measurability [17]. If one lifts the restriction to projective measurements to allow arbitrary POVMs, then one is forced to modify KS-noncontextuality in order to avoid the pathology of trivial POVMs. We do this in a principled way in this paper, building on the approach of Ref. [2].
particular, for Kochen-Specker type experimental scenarios, we will consider the twin notions of preparation noncontextuality and measurement noncontextuality—taken together as a notion of classicality—to obtain noise-robust noncontextuality inequalities that generalize the KS-noncontextuality inequalities of CSW and witness nonclassicality even when the quantum correlations arising from arbitrary quantum measurements on any quantum state are allowed. A key innovation of this approach is that it treats all measurements in an operational theory on an equal footing. No definition of “sharpness” is needed to justify or derive noncontextuality inequalities in this approach. Furthermore, if certain idealizations are presumed about the operational statistics, then these inequalities formally recover the usual Bell-KS inequalities à la CSW. Note that Bell-KS inequalities can be viewed as an instance of the classical marginal problem \cite{25,27,30}, i.e., as constraints on the (marginal) probability distributions over subsets of a set of observables that follow from requiring the existence of global joint probability distribution over the set of all observables. Since the Bell-KS inequalities are only recovered under certain idealizations, but not otherwise, the noise-robust noncontextuality inequalities we obtain cannot in general be viewed as arising from a classical marginal problem. Hence, they cannot be understood within existing frameworks that rely on this (reduction to the classical marginal problem) property to formally unify the treatment of Bell-nonlocality and KS-contextuality \cite{10,11,13}. This is a crucial distinction relative to the usual Bell-KS inequality type witnesses of KS-contextuality.

We now proceed to develop our framework as follows: Section II reviews the Spekkens framework for generalized noncontextuality \cite{3}; Section III introduces a hypergraph framework that shares features of traditional frameworks for KS-contextuality \cite{10,11} but is also augmented (relative to these traditional frameworks) with the ingredients necessary for obtaining noise-robust noncontextuality inequalities; Section IV defines a new hypergraph invariant that we need later on as a crucial new ingredient in our inequalities; and Section V obtains noise-robust noncontextuality inequalities in the framework defined in Section III and using the hypergraph invariant of Section IV, based on the technique proposed in Ref. \cite{5}. Finally, we conclude with some discussion and open questions in Section VI.

II. SPEKKENS FRAMEWORK

We concern ourselves with prepare-and-measure experiments. A schematic of such an experiment is shown in Figure 1 where, for the sake of simplicity, we imagine a single source device that can perform any preparation procedure of interest (rather than a collection of source devices, each implementing a particular preparation procedure) and a single measurement device that can perform any measurement procedure of interest (rather than a collection of measurement devices, each implementing a particular measurement procedure).\footnote{Note that this is just a conceptual abstraction: in particular, the various possible measurement settings on the measurement device may, for example, correspond to incommensurate measurements in quantum theory. The fact that we represent the different measurement settings by choices of knob settings does not mean that it’s physically possible to implement all the measurements represented by \(M \in M\) on a single measurement device; it only means that the experimenter can choose to implement any of the measurements in the set \(M\) in a given run of the prepare-and-measure experiment. The same comment applies to our abstraction of preparation procedures to knob settings and outcomes of a single source device. Also, we take the each of sets \(M\) and \(S\) to include a tomatographically complete set of measurements and preparations, respectively. This tomatographic completeness property is necessary to be able to implement an actual contextuality experiment, see Refs. \cite{2,22,31}.}

The source device has a source setting, \(S \in S\)\footnote{Where \(S\) is just a set of classical labels for different choices of knob settings on the source device.} that can be chosen to prepare a system in an ensemble of possible preparation procedures, \(\{P_{s|S}\}_{s \in V_S}\), according to some probability distribution \(p(s|S)\). This means that the source device has one classical input \(S\) and two outputs: one output is a classical label \(s \in V_S\) identifying the preparation procedure (in the ensemble \(\{p(s|S), P_{s|S}\}_{s \in V_S}\) that is carried out when source outcome \(s\) is observed for source setting \(S\) (this source event

\[ m \in V_M \]
\[ M \in M \]
\[ s \in V_S \]
\[ P_{s|S} \]
\[ S \in S \]

**FIG. 1.** A prepare-and-measure experiment.
is denoted \([s|S]\), and the other output is a system (quantum or otherwise) prepared according to the source event \([s|S]\), i.e., preparation procedure \(P_{[s|S]}\), with probability \(p(s|S)\). Thus, the assemblage of possible ensembles that the source device can prepare can be denoted by \(\{p(s|S), P_{[s|S]}\}_{s\in S, S\in \Sigma}\).

On the other hand, the measurement device has two inputs, one a classical input \(M \in \mathcal{M}\) specifying the choice of measurement setting to be implemented, and the other input receives the system prepared according to preparation procedure \(P_{[s|S]}\) and on which this measurement \(M\) is carried out. The measurement device has one classical output \(m \in V_M\) denoting the outcome of the measurement \(M\) implemented on a system prepared according to \(P_{[s|S]}\), and which occurs with probability \(p(m|M, S, s)\).

We will be interested in the operational joint probability \(p(m, s|M, S) \equiv p(m|S, M, s)p(s|S)\) for this prepare-and-measure experiment for various choices of \(M \in \mathcal{M}, S \in \Sigma\). Note that this operational description takes as primitive the operations carried out in the lab and restricts itself to predicting the probabilities of classical outcomes (i.e., \(m, s\)) given some interventions (i.e., classical inputs, \(M, S\)). To be able to define noncontextuality, the operational theory should admit a notion of operational equivalence, both for sources and for measurements. Two measurement events \([m|M]\) and \([m'|M']\) are said to be operationally equivalent, denoted \([m|M] \simeq [m'|M']\), if there exists no source event that can distinguish them, i.e.,

\[
p(m, s|M, S) = p(m', s|M', S) \quad \forall [s|S], s \in V_S, S \in \Sigma.
\]

Similarly, two source events \([s|S]\) and \([s'|S']\) are said to be operationally equivalent, denoted \([s|S] \simeq [s'|S']\), if there exists no measurement event that can distinguish them, i.e.,

\[
p(m, s|M, S) = p(m, s'|M, S') \quad \forall [m|M], m \in V_M, M \in \mathcal{M}.
\]

In this paper, we will be primarily interested in the operational equivalence between the source settings themselves rather than the source events, i.e., operational equivalence between settings when one coarse-grains over their outcomes. More precisely, two source settings \(S\) and \(S'\) are said to be operationally equivalent, denoted \([\top|S] \simeq [\top|S']\), if no measurement event can distinguish them once all their outcomes are coarse-grained over, i.e.,

\[
\sum_{s \in V_S} p(m, s|M, S) = \sum_{s' \in V_{S'}} p(m, s'|M, S')
\]

\[
\forall [m|M], m \in V_M, M \in \mathcal{M}.
\]

We use the notation \([\top|S]\) to denote coarse-graining over all the outcomes in \(V_S\), i.e., the source event that “at least one of the source outcomes in the set \(V_S\) occurs for source setting \(S\).”

Given the operational description of the experiment in terms of probabilities \(p(m, s|M, S)\), we want to explore the properties of any underlying ontological model for this operational description. Any such ontological model, defined within the ontological models framework \[\text{OM}\], takes as primitive the physical system (rather than operations on it) that passes between the source and measurement devices, i.e., its basic objects are ontic states of the system, denoted \(\lambda \in \Lambda\), that represent intrinsic properties of the physical system. When a preparation procedure \([s|S]\) is carried out, the source device samples from the space of ontic states \(\Lambda\) according to a probability distribution \(\mu(\lambda|S, s) \in [0, 1]\), where \(\sum_{\lambda \in \Lambda} \mu(\lambda|S, s) = 1\), and the joint distribution over \(s\) and \(\lambda\) given \(S\) is given by \(\mu(\lambda, s|S) \equiv \mu(\lambda|S, s)p(s|S)\). On the other hand, when a system in ontic state \(\lambda\) is input to the measurement device with measurement setting \(M \in \mathcal{M}\), the probability distribution over the measurement outcomes is given by \(\xi(m|M, \lambda) \in [0, 1]\), where \(\sum_{m \in V_M} \xi(m|M, \lambda) = 1\). The operational statistics \(p(m, s|M, S)\) results on account of coarse-graining over \(\lambda\), i.e.,

\[
p(m, s|M, S) = \sum_{\lambda \in \Lambda} \xi(m|M, \lambda)\mu(\lambda, s|S).
\]

As such, it is always possible to build an ontological model for an operational theory. It’s only when additional assumptions are imposed on the ontological model that deciding its existence becomes a nontrivial problem. Such additional assumptions must, of course, play an explanatory role to be worth investigating. The assumption we are interested in is noncontextuality and its purpose is to explain the observed operational equivalences in an operational theory. But before we get to noncontextuality, we need to define what a context is:

A context is any distinction between operationally equivalent procedures.

In quantum theory, for example, the preparation basis of the maximally mixed state of a qubit is an example of a preparation context since uniformly mixing the spin up and down eigenstates along any basis leads to the same quantum state. Similarly, when the statistics of a given measurement is inferred by coarse-graining the statistics obtained from measuring it jointly with one or the other measurement (and the two inferences agree), these latter measurements are examples of measurement contexts for the given measurement, e.g., in quantum theory, consider the case of three Hermitian operators \(A, B, C\) such that \([A, B] = [A, C] = 0\) but \([B, C] \neq 0\), so that \(B\) and \(C\) are measurement contexts for \(A\). Here it’s possible to measure \(A\) jointly with \(B\) or with \(C\).
Noncontextuality, motivated by the methodological principle of the identity of indiscernibles [7], is then an inference from the operational description to the ontological description of an experiment. It posits that the operational equivalences are preserved in the ontological model: the reason one cannot distinguish two operationally equivalent procedures is that there is, ontologically, no difference between them. Mathematically, the assumption of measurement noncontextuality entails that

\[ [m|M] \simeq [m'|M'] \Rightarrow \xi(m|M,\lambda) = \xi(m'|M',\lambda) \quad \forall \lambda \in \Lambda, \tag{5} \]

while the assumption of preparation noncontextuality entails that

\[ [s|S] \simeq [s'|S'] \Rightarrow \mu(\lambda, s|S) = \mu(\lambda, s'|S') \quad \forall \lambda \in \Lambda, \]

\[ [\top|S] \simeq [\top|S'] \Rightarrow \mu(\lambda|S) = \mu(\lambda|S') \quad \forall \lambda \in \Lambda. \tag{6} \]

Here we denote \( \mu(\lambda|S) \equiv \sum_{s \in V_s} \mu(\lambda, s|S) \), etc., for simplicity of notation, rather than use the notation \( \mu(\lambda, \top|S) \) for these coarse-grained distributions.

These are the assumptions of noncontextuality – termed universal noncontextuality – that form the basis of our approach to noise-robust noncontextuality inequalities [11]. Note that the traditional notion of KS-noncontextuality entails, besides measurement noncontextuality above, the assumption of outcome-determinism, i.e., for any measurement event \([m|M],\xi(m|M,\lambda) \in \{0,1\} \) for all \( \lambda \in \Lambda \).

III. HYPERGRAPH APPROACH TO KOCHEN-SPECKER SCENARIOS IN THE SPEKKENS FRAMEWORK

We will use the language of hypergraphs and their subgraphs to study Kochen-Specker type experimental scenarios in a framework that allows for operational noncontextuality inequalities à la Spekkens [7]. The presentation here is a hybrid one, discussing features of the CSW framework [9] [10] in the notation of the AALS framework [11], but extending both in ways appropriate for the purpose of this paper. Our goal is to demonstrate how the graph-theoretic invariants of CSW [10] can be repurposed towards obtaining noise-robust noncontextuality inequalities.

We do this in two parts: first, we define a representation of measurement events in the manner of Refs. [10] [11], and then we define a representation of source events in the spirit of Ref. [1].

A. Measurements

The basic object for representing measurements is a hypergraph, \( \Gamma \), with a set of vertices \( V(\Gamma) \) such that each vertex \( v \in V(\Gamma) \) denotes a measurement outcome, and a set of hyperedges \( E(\Gamma) \) such that each hyperedge \( e \in E(\Gamma) \) is a subset of \( V(\Gamma) \) and denotes a measurement consisting of outcomes in \( e \). Here, \( E \subseteq 2^{V(\Gamma)} \) and \( \bigcup_{e \in E(\Gamma)} e = V(\Gamma) \). Such a hypergraph satisfies the definition of a contextuality scenario à la AALS [11]. We will further assume, unless specified otherwise, that the hypergraph is simple: that is, \( e_1, e_2 \in E(\Gamma) \) and \( e_1 \subseteq e_2 \Rightarrow e_1 = e_2 \), or that no hyperedge is a strict subset of another. Such hypergraphs are also called Sperner families [37]. Two measurement events are said to be mutually exclusive if the vertices denoting them appear in a common hyperedge, i.e., if they can be realized in a single measurement. Here, “exclusive” refers to the fact that both measurement events cannot occur together for a given source event when the measurement corresponding to a hyperedge in which they appear together is implemented: hence, the sum of their occurrence probabilities cannot exceed 1.

A probabilistic model on \( \Gamma \) is an assignment of probabilities to the vertices \( v \in V(\Gamma) \) such that \( p(v) \geq 0 \) for all \( v \in V(\Gamma) \) and \( \sum_{v \in V(\Gamma)} p(v) = 1 \) for all \( e \in E(\Gamma) \). Here we are assuming that, in fact, every vertex \( v \) represents an equivalence class of measurement events, denoted \([m|M]\), and every edge \( e \) represents an equivalence class of measurements, denoted \( M \). This assumption is implicit in previous (hyper)graph-theoretic approaches to KS-contextuality [10] [11]. The fact that each \( v \) represents an equivalence class of measurement events, \([m|M]\), means that

1. any probabilistic model on \( \Gamma \), viewed as operational probabilities for a given source event (that is \( p(v) \equiv p(v|S,s) \equiv p(m|M,S,s) \)), respects (by definition) the operational equivalences we have presumed between measurement events in the operational description of the experiment [12] and

2. any probabilistic model on \( \Gamma \), viewed as ontological probabilities for a given ontic state (that is, \( p(v) \equiv p(v|\lambda) \equiv \xi(m|M,\lambda) \)), respects (by definition) the assumption of measurement noncontextuality with respect to the presumed operational equivalences between measurement events.

Thus, it’s the structure of a contextuality scenario that dictates the operational equivalences between measurement events that are of interest and it’s the definition

\[ \sum_{v \in E(\Gamma)} p(v|S,s) = 1 \]

Note that two measurements \( M, M' \) are operationally equivalent if every measurement event of one is operationally equivalent to a distinct measurement event of the other. That is, there is a bijective correspondence (of operational equivalence) between the two sets of measurement events.

12 The fact that a given vertex, say \( v \in V(\Gamma) \), appears in multiple hyperedges, say \( E' \equiv \{ e \in E(\Gamma) | v \in e \} \), means that the measurement events corresponding to this vertex, i.e., \( \{v|e\} \), for all these hyperedges \( e \in E' \), are operationally equivalent, and the equivalence class of these measurement events is denoted by the vertex \( v \) itself. In the case of quantum theory, for example, \( v \) can represent a positive operator that appears in different positive operator-valued measures (POVMs) represented by the hyperedges.
of a probabilistic model on such a contextuality scenario which ensures that such operational equivalences are respected by the operational probabilities as well as the fact that measurement noncontextuality is respected when the probabilistic model is viewed as an ontological assignment of probabilities by a fixed ontic state.

We will therefore often write \( p(m, s|M, S) \) as \( p(v, s|S) \) and \( p(m|M, S, s) \) as \( p(v|S, s) \), where \([s|S]\) is a source event. Similarly, we will also write \( \xi(m|M, \lambda) \) as \( p(v|\lambda) \), where \( \lambda \) is an ontic state.

Orthogonality graph of \( \Gamma \), \( O(\Gamma) \): Given the hypergraph \( \Gamma \), we construct its orthogonality graph \( O(\Gamma) \): that is, the vertices of \( O(\Gamma) \) are given by \( V(O(\Gamma)) = V(\Gamma) \), and the edges of \( O(\Gamma) \) are given by \( E(O(\Gamma)) = \{(v, v')|v, v' \in e \text{ for some } e \in E(\Gamma)\} \). Each edge of \( O(\Gamma) \) denotes the exclusivity of the two measurement events it connects, i.e., the fact that they can occur as outcomes of a single measurement.

For any Bell-KS inequality constraining correlations between measurement events from \( O(\Gamma) \) (when all measurements are implemented on a given source event), we construct a subgraph \( G \) of \( O(\Gamma) \) such that the vertices of \( G \), i.e., \( V(G) \), correspond to measurement events that appear in the inequality with nonzero coefficients, and two vertices share an edge in \( G \) if and only if they share an edge in \( O(\Gamma) \). More explicitly, consider a Bell-KS expression

\[
R([s|S]) = \sum_{v \in V(G)} w_v p(v|S, s),
\]

where \( w_v > 0 \) for all \( v \in V(G) \). A Bell-KS inequality imposes a constraint of the form \( R([s|S]) \leq R_{KS} \), where \( R_{KS} \) is the upper bound on the expression in any operational theory that admits a KS-noncontextual ontological model. Often, but not always, these inequalities are simply of the form where \( w_v = 1 \) for all \( v \in V(G) \). In keeping with the CSW notation \([10]\), we will denote the general situation by a weighted graph \((G, w)\), where \( w \) is a function that maps vertices \( v \in V(G) \) to weights \( w_v > 0 \). See Figures 2 and 3 for an example from the Klyachko-Can-Binicioğlu-Shumovsky (KCBS) scenario \([10, 38]\).

Below, we make some remarks clarifying the scope of the framework described above before we move to the case of sources.

1. Classification of probabilistic models

We classify the probabilistic models on a hypergraph \( \Gamma \) as follows:

- KS-noncontextual probabilistic models, \( C(\Gamma) \): a probabilistic model which is a convex combination of deterministic assignments \( p : V(\Gamma) \rightarrow [0, 1] \), where \( \sum_{v \in e} p(v) = 1 \) for all \( e \in E(\Gamma) \). In Ref. \([11]\), this is referred to as a “classical model” \(^{13}\).

Note that we call \( \Gamma \) KS-colourable if \( C(\Gamma) \neq \emptyset \) and we call it KS-uncolourable if \( C(\Gamma) = \emptyset \).

- Consistent exclusivity satisfying probabilistic models, \( CE(\Gamma) \): a probabilistic model on \( \Gamma \), \( p : V(\Gamma) \rightarrow [0, 1] \), such that (in addition to satisfying the definition of a probabilistic model), \( \sum_{v \in e} p(v) \leq 1 \) for all cliques \( c \) in the orthogonality graph \( O(\Gamma) \). This is the same as the set of \( E1 \) probabilistic models of Ref. \([10]\).

\(^{13}\) We use a different term because we are advocating a revision of the notion of classicality from KS-noncontextuality to generalized noncontextuality à la Spekkens.
Note that a clique in the orthogonality graph $O(\Gamma)$ is a set of vertices that are pairwise exclusive (i.e., every vertex in this set shares an edge with every other vertex).

- General probabilistic models, $G(\Gamma)$: Any $p$ that satisfies the definition of a probabilistic model is a general probabilistic model.

We therefore have

$$C(\Gamma) \subseteq C^1(\Gamma) \subseteq G(\Gamma) \tag{8}$$

for any hypergraph $\Gamma$.

2. Structural Specker’s principle vs. Statistical Specker’s principle

The CSW framework [10] restricts the scope of probabilistic models on a hypergraph to those satisfying consistent exclusivity (the $E1$ probabilistic models), motivated by what is sometimes called Specker’s principle [35]: that is, “if you have several questions and you can answer any two of them, then you can also answer all of them”. If by “questions” we understand measurement settings, then the principle says that a set of pairwise jointly implementable measurement settings is itself jointly implementable. Note that when we say a set of measurement settings is “jointly implementable”, “jointly measurable”, or “compatible”, we mean that there exists another choice of a single measurement setting in the theory such that this measurement setting can reproduce the statistics of all the measurement settings in the set by coarse-graining. See Ref. [29] for an overview of joint measurability in quantum theory. As such, in its application to measurement settings, Specker’s principle is a constraint on the structure of measurements allowed in a physical theory that respects it, e.g., measurement settings that correspond to projective measurements or PVMs (projection valued measures) in quantum theory. This is, for example, the reading adopted in Ref. [40]. On the other hand, we will often also refer to the “joint measurability” of a set of measurement events, by which we mean that this set of measurement events is a subset of the set of measurement outcomes for some choice of measurement setting. At the level of measurement events [14], then, there are two distinct ways to read Specker’s principle that one needs to keep in mind which we distinguish as structural Specker’s principle vs. statistical Specker’s principle. We define these two readings below:

- Structural Specker’s principle imposes a structural constraint on a contextuality scenario $\Gamma$.

This (strong) reading of Specker’s principle applies to any set of measurement events, say $M \subseteq V(\Gamma)$, where every pair of measurement events can arise as outcomes of a single measurement: that is, for each pair $\{v, v'\} \subseteq M$, there exists some $e \in E(\Gamma)$ such that $\{v, v'\} \subseteq e$. The principle then states:

*Given a set $M$ of pairwise jointly measurable measurement events in some contextuality scenario $\Gamma$, all the measurement events in $M$ are jointly measurable, i.e., all the measurement events in the set can arise as outcomes of a single measurement: $M \subseteq e$ for some $e \in E(\Gamma)$.*

Alternatively, the constraint of structural Specker’s principle can be restated as:

*Every clique in the orthogonality graph of $\Gamma$, $O(\Gamma)$, is a subset of some hyperedge in $\Gamma$.*

Note that we haven’t said anything directly about probabilities here: any $\Gamma$ satisfying the above property is said to satisfy structural Specker’s principle.

- Statistical Specker’s principle (or consistent exclusivity) imposes a statistical constraint on probabilistic models on any contextuality scenario $\Gamma$ representing measurement events in an operational theory.

This (weak) reading of Specker’s principle imposes an additional constraint on a probabilistic model $p \in G(\Gamma)$ (thus defining $C^1(\Gamma) \subseteq G(\Gamma)$), namely:

*Given a set $M$ of pairwise jointly measurable measurement events, $p$ satisfies $\sum_{v \in M} p(v) \leq 1$.*

This can also be expressed as:

*A probabilistic model $p \in G(\Gamma)$ is said to satisfy statistical Specker’s principle if the sum of probabilities it assigns to the vertices of every clique in the orthogonality graph of $\Gamma$, $O(\Gamma)$, does not exceed 1, i.e., $\sum_{v \in c} p(v) \leq 1$ for all cliques $c$ in $O(\Gamma)$. All probabilistic models that satisfy this constraint define the set of probabilistic models $C^1(\Gamma)$ (or $E1$ for any contextuality scenario $\Gamma$ regardless of whether $\Gamma$ satisfies structural Specker’s principle. Clearly, $C^1(\Gamma) \subseteq G(\Gamma)$.

Any probabilistic model $p$ on $\Gamma$ such that $p \in C^1(\Gamma)$ is said to satisfy statistical Specker’s principle or, equivalently, consistent exclusivity [11].

Probabilistic models on any hypergraph belonging to the set of $\Gamma$ which satisfy the (strong) structural Specker’s principle obviously satisfy the (weak) statistical Specker’s principle. This holds simply on account of the structure of such $\Gamma$: that is, for all $\Gamma$ satisfying structural Specker’s principle, we have $C^1(\Gamma) = G(\Gamma)$, so that any probabilistic model $p \in G(\Gamma)$ also satisfies statistical Specker’s principle: $p \in C^1(\Gamma)$. To see that $C^1(\Gamma) = G(\Gamma)$ for any $\Gamma$ satisfying structural Specker’s principle, note that every clique $c$ in $O(\Gamma)$ is a subset of some hyperedge in $\Gamma$,
hence for every clique $c \sum_{v \in c} p(v) \leq 1$ for all $p \in G(\Gamma)$, i.e., $p \in C\mathcal{E}^1(\Gamma)$.

While structural Specker’s principle, imposing a constraint on the structure of $\Gamma$, implies that $C\mathcal{E}^1(\Gamma) = G(\Gamma)$\(^{15}\), it remains an open question whether the converse is true:

*That is, given that $C\mathcal{E}^1(\Gamma) = G(\Gamma)$ for some $\Gamma$, is it the case that $\Gamma$ must then necessarily satisfy structural Specker’s principle, namely, that every clique in $O(\Gamma)$ is a subset of some hyperedge in $\Gamma$?*

A positive answer to this question would answer Problem 7.2.3 of Ref. \(^{11}\) asking for a characterization of $\Gamma$ for which $C\mathcal{E}^1(\Gamma) = G(\Gamma)$.

---

3. *What does it mean for an operational theory to satisfy structural/statistical Specker’s principle?*

We have so far defined structural Specker’s principle as a constraint on $\Gamma$ and statistical Specker’s principle as a constraint on a probabilistic model on any $\Gamma$. Any operational theory would typically allow many possible $\Gamma$ to be realized by its measurement events as well as many possible probabilistic models to be realized on any $\Gamma$ representing its measurement events. It will be useful for our discussion to define what we mean when we say an operational theory, say $\mathcal{T}$, satisfies structural Specker’s principle or that it satisfies statistical Specker’s principle.

We denote by $\mathcal{F}(\Gamma)$ the set of probabilistic models achievable on any $\Gamma$ by an operational theory $\mathcal{T}$. Since an operational theory can only put further constraints on probabilistic models in $G(\Gamma)$, we obviously have: $\mathcal{F}(\Gamma) \subseteq G(\Gamma)$.

1. **$\mathcal{T}$ satisfies statistical Specker’s principle:** We say an operational theory $\mathcal{T}$ satisfies statistical Specker’s principle if $\mathcal{F}(\Gamma) \subseteq C\mathcal{E}^1(\Gamma) \subseteq G(\Gamma)$ for all $\Gamma$\(^{16}\).

Since the satisfaction of statistical Specker’s principle is a constraint on the statistical predictions of $\mathcal{T}$, there must be some fact about the structure of $\mathcal{T}$ that leads to this constraint. This fact enforcing statistical Specker’s principle could be some restriction arising from the structure of allowed measurement events and/or even the structure of allowed preparations in the operational theory $\mathcal{T}$. For instance, this is the case for quantum theory when one only considers projective measurements implemented on an arbitrary quantum state, i.e., $\mathcal{O}(\Gamma) \subseteq C\mathcal{E}^1(\Gamma) \subseteq G(\Gamma)$, where $\mathcal{O}(\Gamma)$ denotes the set of probabilistic models that can be obtained in this way. More generally, one could relax the no-restriction hypothesis \(^{11}\) in some particular way in $\mathcal{T}$ so that not all probabilistic models in $G(\Gamma)$ are allowed in $\mathcal{F}(\Gamma)$. In the case of quantum theory, restricting attention to only projective measurements (as we just pointed out) rather than the more general case allowing arbitrary POVMs is one way of restricting the set of possible probabilistic models realizable with quantum states and measurements to a strict subset of $G(\Gamma)$. Allowing arbitrary POVMs would lead to a violation of statistical Specker’s principle by probabilistic models arising from quantum theory.

Let us now define what it means for an operational theory $\mathcal{T}$ to satisfy structural Specker’s principle.

2. **$\mathcal{T}$ satisfies structural Specker’s principle:** An operational theory $\mathcal{T}$ is said to satisfy structural Specker’s principle if for any set of measurement events that are pairwise jointly measurable, i.e., measurement events in each pair arise as outcomes of some measurement in the theory, it is the case that all the measurement events in the set are jointly measurable, i.e., all the measurement events in the set arise as outcomes of a single measurement in the theory.

We will now show that a theory which satisfies structural Specker’s principle also satisfies statistical Specker’s principle. To do this, we consider a contextuality scenario $\Gamma$ which may not satisfy structural Specker’s principle and from it construct a contextuality scenario $\Gamma'$ which does satisfy the principle. The construction proceeds as follows:

1. Construct $O(\Gamma)$.
2. Turn each clique in $O(\Gamma)$ that is a hyperedge in $\Gamma$ to a hyperedge in a new hypergraph $\Gamma'$. That is, $\Gamma'$ is such that $V(\Gamma') \subseteq V(\Gamma')$ and $E(\Gamma') \subseteq E(\Gamma')$.
3. Turn each maximal clique $c$ in $O(\Gamma)$ that is not a hyperedge in $\Gamma$ to a hyperedge in $\Gamma'$ and include an additional vertex $v_c$ in this hyperedge. Here, a maximal clique in a graph is a clique that is not a strict subset of another clique, i.e., there is no vertex outside the clique that shares an edge with each vertex in the clique.

We then have for the hyperedges of $\Gamma'$,

$$E(\Gamma') = E(\Gamma) \cup \{c \cup \{v_c\}\}_{c \in C},$$

where $C$ is the set of maximal cliques in $O(\Gamma)$ that are not hyperedges in $\Gamma$.

Note that as long as a theory $\mathcal{T}$ satisfies structural Specker’s principle, converting maximal cliques in $O(\Gamma)$ that are not hyperedges in $\Gamma$ to hyperedges

---

\(^{15}\) Thus partially answering the open Problem 7.2.3 of Ref. \(^{11}\) and thereby (trivially) implying that statistical Specker’s principle holds for all probabilistic models defined on such $\Gamma$.

\(^{16}\) That is, instead of considering only a particular probabilistic model on a particular $\Gamma$, we now consider the satisfaction of statistical Specker’s principle by a whole set of probabilistic models, namely, $\mathcal{F}(\Gamma)$, for all $\Gamma$.\[^9\]
in \( \Gamma' \) is a valid move within the theory since the resulting hyperedge would indeed constitute a valid measurement in the theory.

If \( C = \emptyset \) (i.e., \( \Gamma \) satisfies structural Specker’s principle), then we just have \( E(\Gamma') = E(\Gamma) \).

4. The resulting contextuality scenario \( \Gamma' \) is thus given by: \( V(\Gamma') = V(\Gamma) \cup \{ v_c \}_{c \in C} \) and \( E(\Gamma') = E(\Gamma) \cup \{ c \cup \{ v_c \}_{c \in C} \} \).

If \( C = \emptyset \) we just have \( V(\Gamma') = V(\Gamma) \) and \( E(\Gamma') = E(\Gamma) \) so that \( \Gamma' = \Gamma \) (i.e., the two hypergraphs are isomorphic).

Our construction of \( \Gamma' \) leads to the following properties:

- \( \Gamma' \) satisfies structural Specker’s principle (by construction) since every clique in \( O(\Gamma') \) is a subset of some hyperedge in \( \Gamma' \). Hence, it’s also the case that statistical Specker’s principle holds for probabilistic models on \( \Gamma' \) as \( CE(\Gamma') = G(\Gamma') \).

Note that the construction of \( \Gamma' \) relied on the fact that the theory we are considering satisfies structural Specker’s principle. If the theory doesn’t satisfy this principle, but one goes ahead with the construction of \( \Gamma' \), then the new hyperedges in \( \Gamma' \) may not constitute valid measurements in the theory.

- Probabilistic models in \( G(\Gamma') \) are in one-to-one correspondence with probabilistic models in \( CE(\Gamma') \): for any probabilistic model \( pr \in CE(\Gamma') \), we can define a probabilistic model \( pr' \in G(\Gamma') \) as \( pr'(v) = pr(v) \) for all \( v \in V(\Gamma) \) and \( pr'(v_c) = 1 - \sum_{v \in v_c} pr(v) \) for all \( c \in C \). Similarly, for any \( pr' \in G(\Gamma') \) we have a probabilistic model \( pr \in CE(\Gamma') \) if we simply ignore the probabilities assigned to the vertices \( v_c \), \( c \in C \), which do not appear in \( \Gamma \).

Hence, the set of probabilistic models on \( \Gamma \) that satisfy statistical Specker’s principle, i.e., \( CE(\Gamma) \), are in one-to-one correspondence with the set of probabilistic models on \( \Gamma' \) which (by construction) satisfies structural Specker’s principle so that \( CE(\Gamma') = G(\Gamma') \).

We therefore have that \( CE(\Gamma) = CE(\Gamma')_{V(\Gamma)} \), where \( CE(\Gamma')_{V(\Gamma)} \) denotes the probabilistic models induced on \( \Gamma \) by those on \( \Gamma' \) (ignoring the probabilities assigned to vertices in \( V(\Gamma') \setminus V(\Gamma) \)).

It is conceivable that a particular \( \Gamma \) may not admit probabilistic models from an operational theory \( \mathbb{T} \), i.e., \( \mathcal{F}(\Gamma) = \emptyset \). On the other hand, if \( \Gamma \) admits a representation in terms of measurement events admissible in \( \mathbb{T} \), so that \( \mathcal{F}(\Gamma) \neq \emptyset \), then two possibilities arise: \( \Gamma \) satisfies structural Specker’s principle or it doesn’t. If \( \Gamma \) satisfies structural Specker’s principle then any probabilistic model in \( \mathcal{F}(\Gamma) \) will satisfy structural Specker’s principle and we have \( \Gamma' = \Gamma \). If \( \Gamma \) does not satisfy structural Specker’s principle, we consider its relation with the contextuality scenario \( \Gamma' \) constructed from it that does satisfy structural Specker’s principle. Such a \( \Gamma' \) admits a representation in a theory \( \mathbb{T} \) satisfying structural Specker’s principle (that is, \( \mathcal{F}(\Gamma') \neq \emptyset \)) as long as \( \Gamma \) admits such a representation (that is, \( \mathcal{F}(\Gamma) \neq \emptyset \)). Indeed, it’s the satisfaction of structural Specker’s principle in \( \mathbb{T} \) that renders the construction of \( \Gamma' \) from \( \Gamma \) physically allowed in \( \mathbb{T} \).

Thus, in a theory \( \mathbb{T} \) that satisfies structural Specker’s principle, the following holds: for every probabilistic model \( pr \in \mathcal{F}(\Gamma) \) (\( \subseteq CE(\Gamma) \)), \( \Gamma' \) admits a corresponding probabilistic model \( pr' \in \mathcal{F}(\Gamma') \) satisfying \( pr'(v) = pr(v) \) for all \( v \in V(\Gamma) \) and \( pr'(v_c) = 1 - \sum_{v \in v_c} pr(v) \) for all \( c \in C \), where \( C \) is the set of maximal cliques in \( O(\Gamma) \) such that none of them is a hyperedge in \( \Gamma \). Similarly, given \( pr' \in \mathcal{F}(\Gamma') \) (\( \subseteq CE(\Gamma') \)), \( pr \in \mathcal{F}(\Gamma) \) is uniquely fixed: it’s obtained by just neglecting the probabilities assigned by \( pr' \) to the vertices in \( V(\Gamma') \setminus V(\Gamma) \).

We must therefore have \( \mathcal{F}(\Gamma) = \mathcal{F}(\Gamma')_{V(\Gamma)} \) for any \( \Gamma \), where \( \mathcal{F}(\Gamma')_{V(\Gamma)} \) denotes the set of probabilistic models induced on \( \Gamma \) by the set of probabilistic models in \( \mathcal{F}(\Gamma') \) under the correspondence we have already established above. We can now state and prove the following theorem:

**Theorem 1.** If an operational theory \( \mathbb{T} \) satisfies structural Specker’s principle, then it also satisfies statistical Specker’s principle.

**Proof.** For any \( \Gamma \) that does not admit a probabilistic model in \( \mathbb{T} \), i.e., \( \mathcal{F}(\Gamma) = \emptyset \), statistical Specker’s principle is trivially satisfied since \( \mathcal{F}(\Gamma) = \emptyset \subseteq CE(\Gamma) \subseteq G(\Gamma) \).

For any \( \Gamma \) that does admit a probabilistic model in \( \mathbb{T} \), i.e., \( \mathcal{F}(\Gamma) \neq \emptyset \), we can have one of two possibilities: either it satisfies structural Specker’s principle, in which case \( \mathcal{F}(\Gamma) \subseteq CE(\Gamma) = G(\Gamma) \), or it doesn’t, in which case we consider the \( \Gamma' \) constructed from it following the recipe we have already outlined so that we have:

\[
\mathcal{F}(\Gamma')_{V(\Gamma)} \subseteq G(\Gamma')_{V(\Gamma)} = CE(\Gamma')_{V(\Gamma)} = CE(\Gamma).
\]

Since \( \mathbb{T} \) satisfies structural Specker’s principle, we have \( \mathcal{F}(\Gamma) = \mathcal{F}(\Gamma')_{V(\Gamma)} \), which immediately implies that \( \mathcal{F}(\Gamma) \subseteq CE(\Gamma) \). That is, the theory \( \mathbb{T} \) satisfies statistical Specker’s principle on \( \Gamma \): \( \mathcal{F}(\Gamma) \subseteq CE(\Gamma) \subseteq G(\Gamma) \).

Overall, we have the desired result: \( \mathbb{T} \) satisfies structural Specker’s principle \( \Rightarrow \mathcal{F}(\Gamma) \subseteq CE(\Gamma) \subseteq G(\Gamma) \) for all \( \Gamma \), i.e., \( \mathbb{T} \) satisfies statistical Specker’s principle. \( \square \)

Thus, one way of enforcing that a particular operational theory \( \mathbb{T} \) satisfies statistical Specker’s principle — that is, \( \mathcal{F}(\Gamma) \subseteq CE(\Gamma) \subseteq G(\Gamma) \) for all \( \Gamma \) — is to require that it satisfies structural Specker’s principle, a constraint on the structure of measurement events in \( \mathbb{T} \). This is, for example, what is achieved in Ref. [31] by invoking a notion of “sharpness” for measurement events in an operational theory such that any set of sharp measurement events that are pairwise jointly measurable are all jointly measurable. That is, structural Specker’s principle is satisfied in a theory with such sharp measurement
events and, consequently, statistical Specker’s principle, or what is more conventionally called consistent exclusivity [11], is also satisfied. But it’s conceivable that there may be other ways to ensure that only a subset of $C^{E_{1}}(\Gamma)$ probabilistic models are allowed in $\mathcal{F}(\Gamma)$ for any $\Gamma$. What we wish to emphasize here is that it is by no means obvious (or at least, it needs to be proven) that the only way to restrict the set of probabilistic models $\mathcal{F}(\Gamma)$ to a subset of $C^{E_{1}}(\Gamma)$ for any $\Gamma$ is to require that the theory $\mathcal{T}$ satisfy structural Specker’s principle [12].

Note that statistical Specker’s principle (or consistent exclusivity) is so intrinsic to the CSW approach [10] that they do not consider probabilistic models that do not satisfy this principle [15]. This will become important when we consider the fact that nonprojective measurements in quantum theory do not satisfy Specker’s principle, structural or statistical (at the level of measurement events), nor even the stronger statement of Specker’s principle for measurement settings (cf. Ref. [40]). Indeed, such measurements admit contextuality scenarios $\Gamma$ that are not possible with projective measurements, such as the one from three binary-outcome POVMs that are pairwise jointly measurable but not treeply so [13]–[16], and the probabilistic models they give rise to can only be accommodated in the most general set of probabilistic models, $\mathcal{G}(\Gamma)$, since trivial POVMs can realize any probabilistic model at all. Specker’s principle, structural or statistical, was motivated by the fact that projective measurements in quantum theory have the property that any set of pairwise orthogonal projectors can be measured together. This principle (in either reading, structural or statistical) would be obeyed in any theory where measurements have this property, and indeed, the more recent attitude of its proponents [28] is to restrict attention to “sharp” measurements in such theories [31, 32], where the definition of “sharp” ensures the property of pairwise jointly measurable projective measurements in such theories [31, 32], where the definition of “sharp” ensures the property of pairwise jointly measurable projective measurements in such theories [31, 32], where the definition of “sharp” ensures the property of pairwise jointly measurable projective measurements in such theories [31, 32], where the definition of “sharp” ensures the property of pairwise jointly measurable projective measurements in such theories [31, 32], where the definition of “sharp” ensures the property of pairwise jointly measurable projective measurements in such theories [31, 32], where the definition of “sharp” ensures the property of pairwise jointly measurable projective measurements in such theories [31, 32].

As we have already noted, a noise-robust noncontextuality inequality of the type in Ref. [1] that is based on a logical proof of the KS theorem is not even obtainable if one restricted attention to probabilistic models satisfying $CE^{1}$. The upper bound on that inequality comes from a probabilistic model that does not satisfy $CE^{1}$. Indeed, any putative theory yielding the set of almost quantum correlations (which satisfy statistical Specker’s principle) [12] cannot satisfy Specker’s principle — that pairwise joint implementable measurement settings are all jointly implementable — for any notion of sharp measurements [30]. Whether structural Specker’s principle, which is defined at the level of measurement events, can be upheld for an almost quantum theory — so that it falls in the category of operational theories with sharp measurements envisaged in Ref. [31] — remains an open question.

4. Remark on the classification of probabilistic models: why we haven’t defined “quantum models” as those obtained from projective measurements

The reader may note that we haven’t tried to define any notion of a “quantum model” so far, having only adopted the definitions of Ref. [11] for KS-noncontextual models ($C(\Gamma)$), for models satisfying consistent exclusivity ($C^{E_{1}}(\Gamma)$), and for general probabilistic models ($\mathcal{G}(\Gamma)$). The reason for this is that we do not wish to restrict ourselves to projective measurements in defining a “quantum model”, unlike the traditional Kochen-Specker approaches [10]–[11]. In Ref. [11], a quantum model is defined as a probabilistic model that can be realized in the following manner: assign projectors $\{\Pi_{e}\}_{e \in V(\Gamma)}$ (defined on any Hilbert space) to all the vertices of $\Gamma$ such that $\sum_{e \in V(\Gamma)} \Pi_{e} = \mathbb{I}$ for all $e \in E(\Gamma)$, and we have $p(v) = \text{Tr}(\rho \Pi_{v})$, for some density operator $\rho$ on the Hilbert space, $\mathbb{I}$ being the identity operator.

On the other hand, allowing arbitrary positive operator-valued measures (POVMs) in a definition of a quantum model (as we would rather prefer) means that, in fact, quantum models on a hypergraph $\Gamma$ are as general
as the general probabilistic models \( G(\Gamma) \), rendering such a definition redundant. This can be seen by noting that for any probabilistic model \( p \in G(\Gamma) \), one can associate positive operators to the vertices of \( \Gamma \) given by \( p(v)I \) such that for any quantum state \( \rho \) on some Hilbert space, we have \( p(v) = \text{Tr}(\rho p(v)I) \), where \( I \) is the identity operator.

Our focus in this paper is not on quantum theory, in particular, even though the need to be able to handle noisy measurements and preparations (particularly, trivial POVMs) in quantum theory can be taken as a motivation for this work. Rather, our focus is on delineating the boundary between operational theories that admit noncontextual ontological models (for Kochen-Specker type experiments, suitably augmented with multiple preparation procedures, as outlined in this paper) and those that don’t by obtaining noise-robust noncontextuality inequalities. In particular, we want these inequalities to indicate the noise thresholds beyond which an experiment cannot rule out the existence of a noncontextual ontological model with respect to the quantities of interest. This also means that making sense of quantum correlations in this approach requires one to pay attention not only to the measurements involved in an experiment but also the preparations; indeed, this shift of focus from measurements alone, to include multiple preparations (or source settings), is a fundamental conceptual difference between our approach and that of traditional Kochen-Specker contextuality frameworks \([10,11,13]\).

5. Scope of this framework

Note that whenever we refer to the “CSW framework”, we mean the framework of Ref. \([10]\), which often differs from the framework of Ref. \([9]\) in some respects, e.g., the normalization of probabilities in a given hyperedge, assumed in \([10]\), but not in \([9]\). In Ref. \([9]\), the authors write:

> Notice that in all of the above we never require that any particular context should be associated to a complete measurement: the conditions only make sure that each context is a subset of outcomes of a measurement and that they are mutually exclusive. Thus, unlike the original KS theorem, it is clear that every context hypergraph \( \Gamma \) has always a classical noncontextual model, besides possibly quantum and generalized models.

On the other hand, in Ref. \([10]\), they write:

> The fact that the sum of probabilities of outcomes of a test is 1 can be used to express these correlations as a positive linear combination of probabilities of events, \( S = \sum w_i P(e_i) \), with \( w_i > 0 \).

The latter presentation \([10]\) is more in line with the “original KS theorem” \([8]\), as well as the presentation in Ref. \([11]\). Since normalization of probabilities is thus presumed in Ref. \([10]\), in keeping with the definition of a probabilistic model we have presented (following \([11]\)), the graph invariants of CSW \([10]\) refer, specifically, to subgraphs \( G \) of those hypergraphs \( \Gamma \) on which the set of KS-noncontextual probabilistic models is nonempty. In particular, our generalization of the CSW framework \([10]\) in this paper says nothing about noise-robust noncontextuality inequalities from logical proofs of the Kochen-Specker theorem \([8]\), which rely on hypergraphs \( \Gamma \) that admit no KS-noncontextual probabilistic models, i.e., KS-uncolourable hypergraphs. It also says nothing for the hypergraphs \( \Gamma \) that do not satisfy the property \( C\mathcal{E}^1(\Gamma) = G(\Gamma) \). An example of such a hypergraph, which is not covered by our generalization of the CSW framework on both counts, is the 18 ray hypergraph first presented in Ref. \([43]\), denoted \( \Gamma_{18} \) (see Fig. 4 and Appendix \([3]\)). Indeed, the study of noise-robust noncontextuality inequalities from such KS-uncolourable hypergraphs was initiated in Ref. \([1]\), and a more exhaustive hypergraph-theoretic treatment of it will be presented in forthcoming work \([44]\). In this paper, we will restrict ourselves to KS-colourable hypergraphs, the study of which was initiated in Ref. \([5]\), and, of these, only those KS-colourable hypergraphs \( \Gamma \) which satisfy \( C\mathcal{E}^1(\Gamma) = G(\Gamma) \). Notice that this is not a limitation of our general approach, which is based on Ref. \([5]\) and applies to any KS-colourable hypergraph, but rather a limitation we inherit from the CSW framework \([10,11]\) since we want to leverage their graph invariants in obtaining our noise-robust noncontextuality inequalities. The study of other KS-

\[\text{FIG. 4. The KS-uncolourable hypergraph from Ref. \([43]\) that is not covered by our generalization of the CSW framework. We denote this hypergraph as } \Gamma_{18}.\]
colourable hypergraphs, in particular those which arise only with nonprojective measurements in quantum theory \cite{14 16} and are outside the scope of traditional frameworks \cite{10 11 13}, will be taken up in future work.

To summarize, the measurement events hypergraphs $\Gamma$ where the present framework (and the CSW framework \cite{10}) applies must satisfy two properties: $C(\Gamma) \neq \emptyset$ (that is, $KS$-colourability) and $CE^1(\Gamma) = \mathcal{G}(\Gamma)$ \footnote{As we have shown, when the operational theory $T$ under consideration satisfies structural Specker’s principle, we can always turn a hypergraph $\Gamma$ that doesn’t satisfy structural Specker’s principle into a hypergraph $\Gamma'$ that satisfies it and for which, therefore, $CE^1(\Gamma') = \mathcal{G}(\Gamma')$ holds. This can be seen as justification for restricting oneself to probabilistic models satisfying consistent exclusivity in the CSW framework \cite{10}; such a restriction is not really a restriction if the theory satisfies structural Specker’s principle. On the other hand, we restrict ourselves to hypergraphs for which $CE^1(\Gamma) = \mathcal{G}(\Gamma)$ without assuming that $T$ satisfies structural Specker’s principle. The justification for this seemingly ad hoc restriction is simply that it is necessary in order to meaningfully leverage the graph invariants of CSW \cite{10} – in particular, the fractional packing number – in our noise-robust noncontextuality inequalities. This will become clear when we obtain our noise-robust noncontextuality inequalities.}

In the next subsection, we define additional notions necessary to obtain noise-robust noncontextuality inequalities that make use of graph invariants from the CSW framework. These notions correspond to source events that are an integral part of our framework.

B. Sources

Having introduced the (hyper)graph-theoretic elements that we need to talk about measurement events, we are now in a position to introduce features of source events that are relevant in the Spekkens framework.

As we have argued previously, we require the measurement events hypergraph $\Gamma$ to be such that $C(\Gamma) \neq \emptyset$ and $CE^1(\Gamma) = \mathcal{G}(\Gamma)$ to be able to obtain noise-robust noncontextuality inequalities that use graph invariants from the CSW framework \cite{10}. Now, in the CSW framework \cite{10}, every Bell-KS expression picks out a particular subgraph $G$ of the orthogonality graph $O(\Gamma)$ of the contextuality scenario $\Gamma$ of interest. The vertices of $G$ denote the measurement events of interest in a given Bell-KS expression and we have the following:

- A general probabilistic model $p \in \mathcal{G}(\Gamma)$ will assign probabilities to vertices in $G$ such that: $p(v) \geq 0$ for all $v \in V(G)$ and $p(v) + p(v') \leq 1$ for every edge $\{v, v'\} \in E(G)$.
- A probabilistic model $p \in CE(\Gamma)$ will assign probabilities to vertices in $G$ such that: $p(v) \geq 0$ for all $v \in V(G)$, $p(v) + p(v') \leq 1$ for every edge $\{v, v'\} \in E(G)$, and

\[
\sum_{v \in c} p(v) \leq 1, \tag{10}
\]

for every clique $c \subseteq V(G)$.

- A probabilistic model $p \in C(\Gamma)$ will assign probabilities to vertices in $G$ such that: $p(v) = \sum_k Pr(k)p_k(v)$, where $Pr(k) \geq 0$, $\sum_k Pr(k) = 1$, and for each $k$, $p_k$ is a set of deterministic assignments $p_k(v) \in \{0, 1\}$ for all $v \in V(G)$, and $p_k(v) + p_k(v') \leq 1$ for every edge $\{v, v'\} \in E(G)$.

Since $\Gamma$ is such that $CE^1(\Gamma) = \mathcal{G}(\Gamma)$, the condition

\[
\sum_{v \in c} p(v) \leq 1 \quad \text{for every clique } c \subseteq V(G)
\]

on the probabilities assigned to vertices in $G$ is redundant. We now define a simplified hypergraph, $\Gamma_G$, obtained from $G$ as follows: convert all maximal cliques in $G$ to hyperedges and add an extra (no-detection) vertex\footnote{“No-detection” in the sense that it denotes the case when none of the measurement events of interest for a given measurement setting occur.} to each such hyperedge.

This $\Gamma_G$, for any $G$, will satisfy the property that $CE^1(\Gamma_G) = \mathcal{G}(\Gamma_G)$ and any probabilistic model on $\Gamma$ assigning probabilities to measurement events in $G$ will correspond to a probabilistic model on $\Gamma_G$ which also assigns the same probabilities to measurement events in $G$. Formally: $V(\Gamma_G) \equiv V(G) \cup \{v_c | c \text{ is a maximal clique in } G\}$, and $E(\Gamma_G) \equiv \{c \cup \{v_c\} | c \text{ is a maximal clique in } G\}$, where $v_c$ is the extra no-detection vertex added to the hyperedge corresponding to maximal clique $c$ in $G$.

We have the following probabilistic model on $\Gamma_G$: given a probabilistic model $p \in \mathcal{G}(\Gamma)$; the probabilities assigned to the vertices in $V(G) \subseteq V(\Gamma_G)$ are the same

\[
\text{FIG. 5. The hypergraph } \Gamma_G \text{ obtained from } G \text{ by adding a no-detection vertex (represented by a hollow circle) to every maximal clique in } G.
\]
as specified by $p \in G(\Gamma)$ and the probabilities assigned to the remaining vertices in $V(G) \setminus V(\Gamma)$ are given by $p(v_e) = 1 - \sum_{v \in e} p(v)$, for every maximal clique $e$ in $G$.

Consider, for example, the KCBS scenario [3, 5, 10, 38]: the 20-vertex $\Gamma$ representing measurement events from five 4-outcome joint measurements (Fig. 2), its 5 vertices $\lambda$ involved in the KCBS inequality (Fig. 3), and 10-vertex hypergraph $\Gamma_G$ constructed from $G$ (Fig. 5).

Given $\Gamma_G$, constructed from $G$, we can now define a hypergraph $\Sigma_G$ of source events as follows: for every hyperedge $e \in E(\Gamma_G)$, corresponding to the choice of measurement setting $M_e$, we define a hyperedge $e \in E(\Sigma_G)$ denoting a corresponding choice of source setting $S_e$. And for every vertex $v \in e(\in E(\Gamma_G))$, we define a vertex $v_e \in e(\in E(\Sigma_G))$. Hence, every measurement event $[v/e]$ in $\Gamma_G$ corresponds to a vertex $v_e$ of $\Sigma_G$, and the number of such vertices in $V(\Sigma_G)$ is $|V(\Gamma_G)||E(\Gamma_G)|$. This means that the operational equivalences between the measurement events that are implicit in $\Gamma_G$ — such as $[v/e]$ is operationally equivalent to $[v/e']$, where $e, e' \in E(\Gamma_G)$ are distinct hyperedges that share the vertex (representing an equivalence class of measurement events) $v \in V(\Gamma_G)$ — are not carried over to the source events, where none is presumed to be operationally equivalent to any other, hence $v_e \in V(\Sigma_G)$ is a different vertex from $v_{e'} \in V(\Sigma_G)$.

Here $v_e \ (v_{e'})$ represents a source event $[s_e | S_e] ([s_{e'} | S_{e'}])$, rather than an equivalence class of source events.

Besides these $|V(\Gamma_G)||E(\Gamma_G)|$ vertices in $V(\Sigma_G)$ and the associated hyperedges $e \in E(\Sigma_G)$, we have an additional hyperedge $e_s \in E(\Sigma_G)$, representing a source setting $S_{e_s}$, containing two new vertices $v^0_{e_s}, v^1_{e_s} \in V(\Sigma_G)$. Here $v^0_{e_s}$ represents the source event $[s_{e_s} = 0 | S_{e_s}]$ and $v^1_{e_s}$ represents the source event $[s_{e_s} = 1 | S_{e_s}]$. Hence, we have $|V(\Sigma_G)| = |V(\Gamma_G)||E(\Gamma_G)| + 2$ and $|E(\Sigma_G)| = |E(\Gamma_G)| + 1$.

The operational equivalence we do presume for $\Sigma_G$ applies to the source settings: all source settings, each represented by coarse grainig the source events in a hyperedge $e \in E(\Sigma_G)$, are operationally equivalent, i.e., $\forall [\lambda | \Sigma_e] \simeq [\lambda | \Sigma_{e'}]$ for all $e, e' \in E(\Sigma_G)$, i.e., $\forall |\lambda | M : p_{m, s_e | M, \Sigma_e} = \sum_{s_{e'}} p_{m, s_{e'} | M, \Sigma_{e'}}$, for all $e, e' \in E(\Sigma_G)$.

An example of such a source events hypergraph was considered in Ref. [1], albeit without the additional source labelled by $e_s$ here [5]. We illustrate it here in Fig. 6 for the KCBS scenario.

\section*{IV. A KEY HYPERGRAPH INVARINT: THE WEIGHTED MAX-PREDICTABILITY}

Without loss of generality, we assume that the extremal probabilistic models on $\Gamma_G$ are in bijective correspondence with the ontic states of the physical system on which the measurements are carried out. Thus, the measurement noncontextual assignment of probabilities given by the response functions $\{\xi(v | \lambda)\}_{v \in V(\Gamma_G)}$ is convexly extremal for all $\lambda \in \Lambda$. The set of extremal probabilistic models on $\Gamma_G$ can then be denoted by the set of ontic states $\Lambda \equiv \Lambda_{\text{det}} \cup \Lambda_{\text{ind}}$, where $\Lambda_{\text{det}}$ corresponds to the set of extremal probabilistic models that assign $\{0, 1\}$-valued probabilities to all the measurement events of $\Gamma_G$, and $\Lambda_{\text{ind}}$ corresponds to the set of extremal probabilistic models that assign $\{0, 1\}$-valued probabilities to some (non-empty subset) of the measurement events of $\Gamma_G$.

We can now define a hypergraph invariant that will be relevant for the operational noncontextuality inequalities we derive:

$$\beta(\Gamma_G, q) \equiv \max_{\lambda \in \Lambda_{\text{ind}}} \sum_{e \in E(\Gamma_G)} q_e \zeta(M_e, \lambda),$$

where $q_e \geq 0$ for all $e \in E(\Gamma_G)$ and $\sum_{e \in E(\Gamma_G)} q_e = 1$ and

$$\zeta(M_e, \lambda) \equiv \max_{v \in e} \xi(v | \lambda)$$

is the maximum probability of occurrence for any outcome of the measurement $M_e$ corresponding to the hyperedge $e \in E(\Gamma_G)$. We call $\beta(\Gamma_G, q)$ the weighted max-predictability of the measurement settings (i.e., hyperedges) in $\Gamma_G$, where the hyperedges $e \in E(\Gamma_G)$ are weighted by the probabilities given by the probability distribution $q \equiv \{q_e\}_{e \in E(\Gamma_G)}$.

\section*{V. NOISE-ROBUST NONCONTEXTUALITY INEQUALITIES}

Consider the positive linear combination of the probabilities of measurement events,

$$R([s | S]) \equiv \sum_{v \in V(G)} w_v p(v | S, s),$$

where $w_v > 0$ for all $v \in V(G)$.
The fundamental result of CSW is that this quantity is bounded for different sets of correlations — KS-noncontextual, those realizable by projective quantum measurements, and those satisfying consistent exclusivity — by graph-theoretic invariants as follows:

\[ \forall[v|s]: R([s|S]) \leq \alpha(G, w) \leq \theta(G, w) \leq \alpha^*(G, w), \]

where KS denotes the set of probabilistic models \( C(\Gamma_G) \), Q denotes the set of probabilistic models on \( \Gamma_G \) achievable by projective quantum measurements, denoted \( Q(\Gamma_G) \), and CE\(^1 \) denotes the set \( CE^1(\Gamma_G) \). The graph invariants of the weighted graph \((G, w)\), namely, \( \alpha(G, w) \), \( \theta(G, w) \), and \( \alpha^*(G, w) \) are defined as follows:

1. Independence number \( \alpha(G, w) \):

\[ \alpha(G, w) = \max_{g \subseteq V(G)} \sum_{v \in g} w_v, \]

where \( g \subseteq V(G) \) is an independent set of vertices of \( G \), i.e., a set of nonadjacent vertices of \( G \), so that none of the vertices in this set shares an edge with any other vertex in the set.

2. Lovasz theta number \( \theta(G, w) \):

\[ \theta(G, w) = \max_{\{\langle u_v \rangle_{v \in V(G)} | \psi \rangle \neq 0 \in V(G)} \sum_{v \in V(G)} w_v |\langle \psi | u_v \rangle|^2, \]

where \( \{\langle u_v \rangle_{v \in V(G)} = \{\langle u_v \rangle_{v \in V(G)} \) (each \( u_v \)) a unit vector in \( \mathbb{R}^d \) is called an orthonormal representation (OR) of the complement of \( G \), namely, \( \bar{G} \), and the unit vector \( |\psi \rangle \in \mathbb{R}^d \) is called a handle. Here \( V(G) = V(G) \) and \( E(G) = \{(v, v') | v, v' \in V(G), (v, v') \not\in E(G)\} \), and we have \( |u_{v'}| = 0 \) for all pairs of nonadjacent vertices, \( (v', v'') \) in \( G \), or equivalently, for all \( (v', v'') \in E(G) \).

3. Fractional packing number \( \alpha^*(G, w) \):

\[ \alpha^*(G, w) = \max_{p_v \in V(G)} \sum_{v \in V(G)} w_v p_v, \]

where \( \{p_v \}_{v \in V(G)} \) is such that \( p_v \geq 0 \) for all \( v \in V(G) \) and \( \sum_{v \in c} p_v \leq 1 \) for all cliques \( c \) in \( G \).

Note that since we are always considering \( \Gamma_G \) such that \( CE^1(\Gamma_G) = \bar{G}(\Gamma_G) \), we, in fact, have the bounds

\[ \forall[v|s]: R([s|S]) \leq R_{KS}(G, w) \leq R_{Q}(\Gamma_G) \leq \alpha(G, w) \leq \alpha^*(G, w), \]

where GPT denotes the full set of probabilistic models on \( \Gamma_G \), i.e., \( \bar{G}(\Gamma_G) \).

In terms of the notation we have already introduced, where \( R([s|S]) \leq R_{KS} \) was a Bell-KS inequality, we now have — from CSW [10] — that \( R_{KS} = \alpha(G, w) \).

We need to define a new quantity not in the CSW framework, namely,

\[ Corr = \sum_{e \in E(\Gamma_G)} q_e \sum_{m_e,s_e} \delta_{m_e,s_e} p(m_e,s_e|M_e,S_e), \]

where \( q_e \in E(\Gamma_G) \) is a probability distribution, i.e., \( q_e \geq 0 \) for all \( e \in E(\Gamma_G) \) and \( \sum_{e \in E(\Gamma_G)} q_e = 1 \), such that \( \beta(\Gamma_G, q) < 1 \) holds. In previous work [11, 12], we have taken \( q \) to be the uniform distribution \( q_e = 1/E(\Gamma_G) \), but the derivation of the noncontextuality inequalities is independent of that choice (as we’ll see here). Also, note that we have chosen a labelling convention for outcomes of source setting \( S_e \) (namely, \( s_e \)) and measurement setting \( M_e \) (namely, \( m_e \)) outcomes corresponding to the measurement event \( [v|e] \) (denoting \([m_e|M_e]) \) and its corresponding source event \( v_e \) (denoting \([s_e|S_e]) \) are both denoted by the same label. An example of this from Figs. 5 and 6 would be to, say, denote the outcomes of a particular \( e \in E(\Gamma_G) \) (measurement setting \( M_e \)) by \( m_e \in \{0, 1, 2\} \) and corresponding outcomes of \( e \in E(\Gamma_G) \) (source setting \( S_e \)) by \( s_e \in \{0, 1, 2\} \); so if \([v|e] \) denotes \([m_e = 0|M_e]) \), then \( v_e \) will denote \([s_e = 0|S_e]) \), etc.

In the ontological model,

\[ R([s|S]) = \sum_{\lambda \in \Lambda} \sum_{\lambda \in \Lambda} w_j p(\lambda) \mu(\lambda|S, s). \]

Defining \( R(\lambda) = \sum_{e \in E(\Gamma_G)} w_j p(\lambda) \), we have that

\[ R([s|S]) = \sum_{\lambda \in \Lambda} R(\lambda) \mu(\lambda|S, s). \]

Similarly,

\[ Corr = \sum_{\lambda \in \Lambda} \sum_{e \in E(\Gamma_G)} q_e \sum_{m_e,s_e} \delta_{m_e,s_e} \xi(m_e|M_e, \lambda) \mu(s_e|S_e, \lambda) \nu(\lambda) \]

\[ = \sum_{\lambda \in \Lambda} Corr(\lambda) \nu(\lambda), \]

where we have used preparation noncontextuality:

\[ \forall e, e' \in E(\Sigma_S) : [|T|S_e] ~ [|T|S_{e'}] \implies \mu(\lambda|S_e) = \mu(\lambda|S_{e'}) \equiv \nu(\lambda), \forall \lambda \in \Lambda. \]

Using the fact that

\[ \nu(\lambda) = \mu(\lambda|S) = \sum_s \mu(\lambda|S, s)p(s|S), \]

for any \( S \equiv S_e, e \in E(\Sigma_S) \), we have

\[ Corr = \sum_s \left( \sum_{\lambda} Corr(\lambda) \mu(\lambda|S, s) \right) p(s|S) \]

\[ = \sum_{s} Corr_s p(s|S). \]

Indeed, for the strongest possible constraint on Corr, one must pick \( q \) such that \( \beta(\Gamma_G, q) \) is minimized.
KS-contextuality is witnessed when for some choice of $|S|$, say $|s_e = 0 S_e|$, we have 
$R(|s_e = 0 S_e|) > R_{KS}$: this means that for some set of $\lambda \in \text{Supp}(\mu, S_e, s_e = 0) \equiv \{ \lambda \in \Lambda : \mu(\lambda|S_e, s_e = 0) > 0 \}$, we have $R(\lambda) > R_{KS}$ and for this set of $\lambda$ one must have $\text{Corr}(\lambda) < 1$ (because these $\lambda \in \Lambda_{\text{ind}}$), which in turn implies that $\text{Corr}_{s_e = 0} \equiv f(R - R_{KS}) < 1$, where $R \equiv R(|s_e = 0 S_e|)$ and the upper bound is a function of the violation of $R \leq R_{KS}$. On the other hand, for $s_e = 1$, we have no constraints: $\text{Corr}_{s_e = 1} \leq 1$. Thus,

$$\text{Corr} = \text{Corr}_{s_e = 0} p(s_e = 0|S_e) + \text{Corr}_{s_e = 1} p(s_e = 1|S_e) \leq p_0 f(R - R_{KS}) + 1 - p_0, \quad (25)$$

where $p_0 \equiv p(s_e = 0|S_e)$. Note that, for any $\lambda \in \Lambda$, $\text{Corr}(\lambda)$ is upper bounded as follows:

$$\text{Corr}(\lambda) \equiv \sum_{e \in E(G)} q_e \sum_{m_e|s_e} \delta_{m_e|s_e} \xi(m_e|M_e, \lambda) \mu(s_e|S_e, \lambda) \leq \sum_{e \in E(G)} q_e \xi(M_e, \lambda) \sum_{s_e} \mu(s_e|S_e, \lambda) = \sum_{e \in E(G)} q_e \xi(M_e, \lambda), \quad (26)$$

where $\xi(M_e, \lambda) \equiv \max_{m_e} \xi(m_e|M_e, \lambda)$. If $\lambda \in \Lambda_{\text{det}}$, then this upper bound is trivial, i.e., $\text{Corr}(\lambda) \leq 1$. On the other hand, for all $\lambda \in \Lambda_{\text{ind}}$, we have

$$\text{Corr}(\lambda) \leq \beta(\Gamma_G, q). \quad (27)$$

Similarly, for $\lambda \in \Lambda_{\text{det}}$ we have $R(\lambda) \leq \alpha(G, w)$, while for $\lambda \in \Lambda_{\text{ind}}$ we have $R(\lambda) \leq \alpha^*(G, w)$.

Defining $\mu_{\text{det}} \equiv \sum_{\lambda \in \Lambda_{\text{det}}} \mu(\lambda|S_e, s_e = 0)$ and $\mu_{\text{ind}} \equiv \sum_{\lambda \in \Lambda_{\text{ind}}} \mu(\lambda|S_e, s_e = 0)$, we now have

$$\mu_{\text{det}} + \mu_{\text{ind}} = 1, \quad (28)$$

$$\text{Corr}_{s_e = 0} \leq \mu_{\text{det}} + \beta(\Gamma_G, q) \mu_{\text{ind}}, \quad (29)$$

$$R \leq \alpha(G, w) \mu_{\text{det}} + \alpha^*(G, w) \mu_{\text{ind}}. \quad (30)$$

Note that assuming $\mu_{\text{det}} = 1$ would reduce these constraints to a standard Bell-KS inequality, $R \leq \alpha(G, w)$. However, since we are not assuming this, simply eliminating $\mu_{\text{det}}$ and $\mu_{\text{ind}}$ from these constraints leads us to

$$\text{Corr}_{s_e = 0} \leq 1 - (1 - \beta(\Gamma_G, q)) \frac{R - \alpha(G, w)}{\alpha^*(G, w) - \alpha(G, w)} \equiv f(R - \alpha(G, w)), \quad (31)$$

where $f(R - \alpha(G, w)) < 1$ if and only if $\beta(\Gamma_G, q) < 1$ and $R - \alpha(G, w) > 0$.

If we are given that $\beta(\Gamma_G, q) < 1$, then we have a trivial upper bound on $\text{Corr}_{s_e = 0}$ for the remaining cases: $f(R - \alpha(G, w)) = 1$ for $R = \alpha(G, w)$ and $f(R - \alpha(G, w)) > 1$ for $R < \alpha(G, w)$.

Thus, our noise-robust noncontextuality inequality now reads:

$$\text{Corr} \leq 1 - p_0 (1 - \beta(\Gamma_G, q)) \frac{R - \alpha(G, w)}{\alpha^*(G, w) - \alpha(G, w)}. \quad (32)$$

which can be rewritten as

$$R \leq \alpha(G, w) + \frac{\alpha^*(G, w) - \alpha(G, w)}{p_0} (1 - \text{Corr}) - \frac{1}{\beta(\Gamma_G, q)}. \quad (33)$$

For a nontrivial upper bound and hence, the possibility of witnessing contextuality via this inequality – the upper bound on $R$ should be strictly bounded above by 1, and the upper bound on $R$ should be strictly bounded above by $\alpha^*(G, w)$ (the algebraic upper bound on $R$), that is

$$p_0 > 0 \text{ and } \beta(\Gamma_G, q) < 1, \quad R > \alpha(G, w), \quad \text{Corr} > 1 - p_0 (1 - \beta(\Gamma_G, q)). \quad (34)$$

These are the minimal benchmarks necessary—besides the requirement of tomographic completeness of a finite set of procedures and the possibility of inferring secondary procedures with exact operational equivalences using convexity of the operational theory [2]—to witness contextuality in a Kochen-Specker type experiment adapted to our framework following Spekkens [4].

Suppose one achieves, by some means, a value of $R = \theta(G, w)$. When would this value be an evidence of contextuality? For this to be the case, we must have:

$$\text{Corr} > 1 - p_0 (1 - \beta(\Gamma_G, q)) \frac{\theta(G, w) - \alpha(G, w)}{\alpha^*(G, w) - \alpha(G, w)}, \quad (35)$$

Now, for the ideal quantum realization where measurement events are projectors, and the corresponding source events are eigenstates, it is always the case that $\text{Corr} = 1$, hence contextuality is witnessed. However, it’s possible to witness contextuality even if $\text{Corr} < 1$, as long as it exceeds the lower bound we specified above. In a sense, for quantum theory, this allows for a quantitative account of the effect of nonprojectiveness in the measurements (or mixedness in preparations) on the possibility of witnessing contextuality, a feature that is absent in traditional Kochen-Specker approaches [9,11,13]. Indeed, as long as one achieves any value of $R > \alpha(G, w)$, it is possible to witness contextuality for a sufficiently high value of $\text{Corr}$ (see Eq. [32]).

A. Example: KCBS scenario

We will now illustrate our hypergraph framework by applying it to the KCBS scenario to make differences with respect to the CSW graph-theoretic framework [10] explicit.
The graph $G$ for the KCBS scenario is given in Fig. 3, the measurement events hypergraph $\Gamma_G$ is given in Fig. 5, and the source events hypergraph $\Sigma_G$ is given in Fig. 6. We then have

$$R([s|S]) = \sum_{v \in V(G)} p(v|S, s), \quad (36)$$

where the (vertex) weights $w_v = 1$ for all $v \in V(G)$, i.e., it’s an unweighted graph and we will use $\alpha(G)$ and $\alpha^*(G)$ to denote its independence number and the fractional packing number, respectively. These are given by

$$\alpha(G) = 2 \quad \text{and} \quad \alpha^*(G) = 5/2. \quad (37)$$

The source-measurement correlation term is given by

$$\text{Corr} = \sum_{e \in E(\Gamma_G)} q_e \sum_{m_e, s_e} \delta_{m_e, s_e} p(m_e, s_e | M_e, S_e) \quad (38)$$

for any choice of probability distribution $q \equiv \{q_e\}_{e \in E(\Gamma_G)}$. Note that the only extremal probabilistic model on $\Gamma_G$ corresponding to an indeterministic assignment (in $\Lambda_{\text{ind}}$) assigns $\xi(v|\lambda) = \frac{1}{2}$ for all $v \in V(G)$. This means

$$\beta(\Gamma_G, q) = \frac{1}{2} \quad \forall q. \quad (39)$$

The noncontextuality inequality of Eq. (38) becomes (in the KCBS scenario)

$$R \leq 2 + \frac{1/2 - \text{Corr}}{p_0}, \quad (41)$$

or

$$R \leq 2 + \frac{1 - \text{Corr}}{p_0}. \quad (42)$$

Recall that the KCBS inequality [10, 38] reads $R \leq 2$ and it would be a valid noncontextuality inequality in our framework if and only if one can find measurements and preparations such that $\text{Corr} = 1$. In the standard KCBS construction [38] that violates the inequality $R \leq 2$, we have the five vertices in $G$ (say $v_i$, $i \in \{1, 2, 3, 4, 5\}$, labelled cyclically) associated with five projectors $\Pi_i = |l_i\rangle\langle l_i|$, $i \in \{1, 2, 3, 4, 5\}$, on a qutrit Hilbert space, given by the vectors $|l_i\rangle = (\sin \theta \cos \phi_i, \sin \theta \sin \phi_i, \cos \theta)$, $\phi_i = \frac{2\pi i}{5}$, and $\cos \theta = \frac{1}{\sqrt{3}}$. The special source event $|s_{e_i} = 0|S_{e_i}\rangle$ is associated with the quantum state $|\psi\rangle = (0, 0, 1)$, so that

$$R = \sum_{i=1}^{5} |\langle l_i|\psi\rangle|^2 = \sqrt{5} > 2. \quad (43)$$

See Fig. 7 for a depiction of the geometric configuration of these vectors.

To turn this KCBS construction into an argument against noncontextuality in our approach, we need additional ingredients beyond the graph $G$. Firstly, for both the measurement events hypergraph $\Gamma_G$ and the source events hypergraph $\Sigma_G$, we denote the hyperedges by $e_i$, $i \in \{1, 2, 3, 4, 5\}$. In $\Gamma_G$, the measurement events for the setting $M_{e_i}$ are given by \{\[m_{e_i} = 0|M_{e_i}\] = $|l_i\rangle\langle l_i|, \[m_{e_i} = 1|M_{e_i}\] = $1 - |l_i\rangle\langle l_i| - |l_{i+1}\rangle\langle l_{i+1}|$, $|m_{e_i} = 2|M_{e_i}\] = $|l_{i+1}\rangle\langle l_{i+1}|\}$, where for $i = 5$, $i + 1 = 1$ (addition modulo 5). Similarly, in $\Sigma_G$, the source events corresponding to source setting $S_{e_i}$ are given by \{\[s_{e_i} = 0|S_{e_i}\] = $|e_i\rangle\langle e_i|, \[s_{e_i} = 2|S_{e_i}\] = $|e_{i+1}\rangle\langle e_{i+1}|$, and \[s_{e_i} = 1|S_{e_i}\] = $1 - |e_i\rangle\langle e_i| - |e_{i+1}\rangle\langle e_{i+1}|\}$, where $p(s_{e_i} = b|S_{e_i}) = \frac{1}{3}$ for all $b \in \{0, 1, 2\}$. The special source setting $S_{e_i}$ consists of source events \{\[s_{e_i} = 0|S_{e_i}\] = $|\psi\rangle\langle \psi|, \[s_{e_i} = 1|S_{e_i}\] = $1 - |\psi\rangle\langle \psi|\}$, where $p(s_{e_i} = 0|S_{e_i}) = \frac{1}{3}$ and $p(s_{e_i} = 1|S_{e_i}) = \frac{2}{3}$. We thus have the operational equivalences we need between the source settings:

$$[\Gamma|S_e] \simeq [\Gamma|S_{e'}] = \frac{1}{3}; \quad \forall e, e' \in E(\Sigma_G). \quad (44)$$

This choice of representation for $\Gamma_G$ and $\Sigma_G$ yields $p_0 = \frac{1}{3}$, $\text{Corr} = 1$, and $R([s_{e_i} = 0|S_{e_i}|) = \sqrt{5}$, so that the inequality

$$R \leq 2 + \frac{1 - \text{Corr}}{p_0}. \quad (45)$$

is violated. However, note that this is an idealization (under which $\text{Corr} = 1$) and, typically, the source events and measurement events will not be perfectly correlated ($\text{Corr} < 1$) and the operational equivalences between the source settings need not correspond to the maximally mixed state. All that is required for a test of noncontextuality using this inequality is that the operational equivalences hold for some choice of preparations and measurements which need not be the same as that in the ideal KCBS construction.
VI. DISCUSSION

A. Measurement-measurement correlations vs. source-measurement correlations

Note that the usual Kochen-Specker experiment, as conceptualized in Refs. 9 [11] [13], for example, involves only the quantity \( R([s|S]) \), representing correlations between various measurement events when all the measurements are implemented on a system prepared according to the same preparation procedure, denoted by the source event \([s|S]\). Thus, \( R \) represents measurement-measurement correlations on a system prepared according to a fixed choice of preparation procedure.

On the other hand, the experiment we have conceptualized in this paper involves, besides the quantity \( R \), a quantity \( C \), representing measurements in terms of their corresponding measurements and characterizing the quality of the measurements in terms of their response to corresponding preparations.

Our noncontextuality inequalities represent a trade-off relation that must hold between \( R \) and \( C \) in an operational theory that admits a noncontextual ontological model. Here we note that the first example of such a tradeoff relation, albeit only for the case of operational quantum theory with unsharp measurements, appeared in Ref. 14 as the Liang-Spekkens-Wiseman (LSW) inequality [15] which has been shown to be experimentally violated in Ref. 16 [17] And, indeed, the developments reported in Ref. [5] and the present paper have their origins in the idea of such a trade-off relation that first appeared in Ref. [14].

B. Can our noise-robust noncontextuality inequalities be saturated by a noncontextual ontological model?

A natural question concerns the tightness of these noncontextuality inequalities, i.e., can they be satisfied by a noncontextual ontological model? This requires one to specify a noncontextual ontological model for which

\[
\text{Corr} = 1 - p_0(1 - \beta(G, q)) \frac{R - \alpha(G, w)}{\alpha^*(G, w) - \alpha(G, w)},
\]

or, equivalently,

\[
R = \alpha(G, w) + \frac{\alpha^*(G, w) - \alpha(G, w)}{p_0} \frac{1 - \text{Corr}}{1 - \beta(G, q)}.
\]

The assumption of measurement noncontextuality is already implicit in our characterization of the response functions \( \xi(m_e|M_e, \lambda) \), and for this reason it is, indeed, trivial to satisfy measurement noncontextuality while saturating these noncontextuality inequalities. Measurement noncontextuality, alone, in fact even allows \( R = \alpha^*(G, w) \). On the other hand, as in traditional Bell-KS type treatments, if outcome determinism is presumed, then we know that there exists a necessary and sufficient set of Bell-KS inequalities (each corresponding to a particular choice of \( R([s|S]) \)) that are saturated by a KS-noncontextual ontological model: this just corresponds to the case \( R([s|S]) = \alpha(G, w) \) for any such Bell-KS inequality. Indeed, our noise-robust noncontextuality inequalities corresponding to these choices of \( R([s|S]) \) can always be saturated when \( \text{Corr} = 1 \), because in that case outcome determinism is justified by preparation noncontextuality (cf. Ref. [5]) and the inequalities are identical to the Bell-KS inequalities.

Since we are not assuming outcome determinism, nor the idealization of \( \text{Corr} = 1 \), what is at stake here is the assumption of preparation noncontextuality. It must be satisfied while saturating the upper bound \((<1)\) in the noncontextuality inequality. This is the critical assumption that any measurement noncontextual model must uphold for it to be universally noncontextual. That is, we have to specify distributions \( \mu(s_e|S_e, \lambda) \) and \( \nu(\lambda) \) such that

\[
\forall \lambda \in \Lambda : \mu(\lambda|S_e) = \nu(\lambda), \forall e \in E(S_G),
\]

and we have

\[
(\alpha^*(G, w) - \alpha(G, w))\text{Corr} + p_0(1 - \beta(G, q))R = (\alpha^*(G, w) - \alpha(G, w)) + p_0\alpha(G, w)(1 - \beta(G, q)),
\]

where

\[
\text{Corr} = \sum_{s_e} p(s_e|S_e, \lambda) \text{Corr}_{s_e},
\]

\[
\text{Corr}_{s_e} = \sum_{\lambda \in \Lambda} \text{Corr}(\lambda) \mu(\lambda|S_e, s_e),
\]

and

\[
R = \sum_{\lambda \in \Lambda} R(\lambda) \mu(s_e = 0|S_e, \lambda).
\]

To show that the noncontextuality inequalities can be saturated, we construct a noncontextual ontological model with the following constraints:

1. For any \( \lambda \in \Lambda_{det} : \)

\[
\forall s_e, e \in E(S_G) \setminus \{e_e\} : \mu(s_e|S_e, \lambda) = \delta_{s_e, x_e},
\]

where \( x_e \) is such that \( \xi(m_e|M_e, \lambda) = \delta_{m_e, x_e} \). This choice of \( \mu(s_e|S_e, \lambda) \) for \( \lambda \in \Lambda_{det} \) ensures that

\[
\text{Corr}(\lambda) = 1 \quad \forall \lambda \in \Lambda_{det}.
\]
2. For any $\lambda \in \Lambda_{\text{ind}}$:

$$\forall s,e \in E(\Sigma G) \setminus \{e_s\} : \mu(s,e|S,e,\lambda) = \delta_{s,e},$$  

where $y_e$ is such that $\xi(m_e) = y_e|M,e,\lambda = \max_{m_e} \xi(m_e|M,e,\lambda)$. This choice of $\mu(s,e|S,e,\lambda)$ for $\lambda \in \Lambda_{\text{ind}}$ ensures that

$$\text{Corr}(\lambda) = \text{Corr}_{\text{ind}}^\lambda \equiv \sum_{e \in E(\Gamma G)} q_e \zeta(M,e,\lambda), \forall \lambda \in \Lambda_{\text{ind}}.$$

3. Let us define $\Lambda_{\text{corr}}^\max \subseteq \Lambda_{\text{ind}}$ and $\Lambda_{\text{max}}^R \subseteq \Lambda_{\text{ind}}$ such that

$$\forall \lambda \in \Lambda_{\text{corr}}^\max : \text{Corr}(\lambda) = \max_{\lambda' \in \Lambda_{\text{ind}}} \text{Corr}_{\text{ind}}^{\lambda'} \equiv \beta(G,q),$$

and

$$\forall \lambda \in \Lambda_{\text{max}}^R : R(\lambda) = \max_{\lambda' \in \Lambda_{\text{ind}}} R(\lambda') \equiv \alpha^*(G,w).$$

For our construction to work, we require that the polytope of probabilistic models on $\Gamma_G$ be such that $\Lambda_{\text{corr}}^\max \cap \Lambda_{\text{max}}^R \neq \emptyset$, i.e., there exist indeterministic extremal probabilistic models corresponding to $\lambda \in \Lambda_{\text{corr}}^\max \cap \Lambda_{\text{max}}^R$ that maximize both $R(\lambda)$ and $\text{Corr}(\lambda)$.

4. The distribution $\mu(\lambda|S,e,=0)$ is such that

$$\sum_{\lambda \in \Lambda_{\text{det}}: R(\lambda) = \alpha(G,w)} \mu(\lambda|S,e,=0)$$

$$= \sum_{\lambda \in \Lambda_{\text{det}}} \mu(\lambda|S,e,=0)$$

$$\equiv \mu_{\text{det}}$$

and

$$\sum_{\lambda \in \Lambda_{\text{corr}}^\max \cap \Lambda_{\text{max}}^R} \mu(\lambda|S,e,=0)$$

$$= \sum_{\lambda \in \Lambda_{\text{ind}}} \mu(\lambda|S,e,=0)$$

$$\equiv \mu_{\text{ind}}.$$

5. The distribution $\mu(\lambda|S,e,=1)$ is such that

$$\sum_{\lambda \in \Lambda_{\text{det}}} \mu(\lambda|S,e,=1) = 1.$$

With these constraints on the noncontextual ontological model in hand, we have:

$$\text{Corr}_{s,e=0} = \mu_{\text{det}} + \mu_{\text{ind}} \beta(G,q)$$

using constraints 1, 2, 3, and 4,

$$\text{Corr}_{s,e=1} = 1$$

using constraint 5. We therefore have

$$\text{Corr} = p_0 \text{Corr}_{s,e=0} + 1 - p_0 = 1 - p_0 \mu_{\text{ind}}(1 - \beta(G,q)).$$

Again, using constraints 1, 2, 3, and 4, we have

$$R = \mu_{\text{det}} \alpha(G,w) + \mu_{\text{ind}} \alpha^*(G,w)$$

$$= \alpha(G,w) + \mu_{\text{ind}} \alpha^*(G,w) - \alpha(G,w).$$

Plugging these expressions for $\text{Corr}$ and $R$ on the left-hand-side of the condition for saturation, Eq. (32), of the noncontextuality inequality, we have

$$(\alpha^*(G,w) - \alpha(G,w)) \text{Corr} + p_0(1 - \beta(G,q)) R$$

$$= p_0(1 - \beta(G,q)) (\alpha(G,w) - \mu_{\text{ind}}(1 - \beta(G,q)))$$

$$+ \mu_{\text{ind}}(\alpha^*(G,w) - \alpha(G,w))$$

$$= \alpha(G,w) - \alpha(G,w) + p_0 \alpha(G,w)(1 - \beta(G,q)),$$

so that our noncontextual ontological model, subject to the specified constraints, saturates the noncontextuality inequality.

The crucial condition we need for this noncontextual ontological model to work is that there exists at least one extremal probabilistic model on $\Gamma_G$ (corresponding to a $\lambda \in \Lambda_{\text{ind}}$) such that $R(\lambda) = \alpha^*(G,w)$ and $\text{Corr}(\lambda) = \beta(G,q)$. For all such $\Gamma_G$, we have shown that our noncontextuality inequalities will be saturated.

This leaves us with some open questions: Does this crucial condition hold for all $\Gamma_G$ of interest? If it doesn’t hold for some $\Gamma_G$, is it still possible: 1) to obtain a different noncontextual ontological model that saturates Eq. (32), or 2) to derive a (tight) noncontextuality inequality that is possibly different from Eq. (32) but is saturated by a noncontextual ontological model?

Note that for the case of $\Gamma_G$ that admit indeterministic extremal probabilistic models with probability assignments only in $\{0, 1\}$, e.g., the $n$-cycle scenarios discussed in Ref. [5], this condition always holds.

C. Can trivial POVMs ever violate these noncontextuality inequalities?

No.

Recall that a trivial POVM is defined as an assignment of positive operators $p(v)I$ to the vertices of $\Gamma_G$, where $I$ is the identity operator on some Hilbert space and $p : V(\Gamma_G) \to [0,1]$, such that $\sum_{v \in e} p(v) = 1$ for all $e \in E(\Gamma_G)$, is a probabilistic model on $\Gamma_G$. 
Consider trivial POVMs corresponding to any KS-noncontextual probabilistic model (that is a convex mixture of deterministic vertices, $\Lambda_{\text{det}}$). The largest value Corr can take in this case is less than or equal to 1. This means that the upper bound on $R$ from our noncontextuality inequality, Eq. [33], will be greater than or equal to $\alpha(G,w)$, whereas we know that for a KS-noncontextual probabilistic model, $R \leq \alpha(G,w)$. Hence, there is no violation of our noncontextuality inequality for such trivial POVMs.

Now consider trivial POVMs that correspond to the indeterministic vertices, $\Lambda_{\text{ind}}$, or their convex mixtures. We know that for these trivial POVMs, Corr can take in this case is less than or equal to 1. This is again trivially satisfied since it reduces to

$$
\beta(\Gamma_G,q) \leq 1 - p_0 (1 - \beta(\Gamma_G,q)) \frac{R - \alpha(G,w)}{\alpha^*(G,w) - \alpha(G,w)}.
$$

(68)

which reduces, for $R > \alpha(G,w)$, to

$$
p_0 \leq \frac{\alpha^*(G,w) - \alpha(G,w)}{R - \alpha(G,w)},
$$

(69)

where the upper bound is greater than or equal to 1, since $\alpha(G,w) < R \leq \alpha^*(G,w)$. This is trivially satisfied since $p_0 \leq 1$.

For $R < \alpha(G,w)$, the sufficient condition of Eq. (68) is again trivially satisfied since it reduces to

$$
p_0 \geq \frac{-\alpha^*(G,w) - \alpha(G,w)}{\alpha(G,w) - R},
$$

(70)

and we must anyway have $p_0 \geq 0$.

For $R = \alpha(G,w)$, the sufficient condition reduces to $\beta(\Gamma_G,q) \leq 1$, which is again trivially satisfied since $\beta(\Gamma_G,q) < 1$ by definition.

In general, a probabilistic model achieved by trivial POVMs can be in the convex hull of both deterministic ($\Lambda_{\text{det}}$) and indeterministic ($\Lambda_{\text{ind}}$) vertices, with the total weight on deterministic vertices denoted by $\Pr(\Lambda_{\text{det}})$ and that on indeterministic vertices by $\Pr(\Lambda_{\text{ind}})$, so that $\Pr(\Lambda_{\text{det}}) + \Pr(\Lambda_{\text{ind}}) = 1$. We then have

$$
\text{Corr} \leq \Pr(\Lambda_{\text{det}}) + \Pr(\Lambda_{\text{ind}})\beta(\Gamma_G,q),
$$

$$
R \leq \Pr(\Lambda_{\text{det}})\alpha(G,w) + \Pr(\Lambda_{\text{ind}})\alpha^*(G,w).
$$

(71)

A sufficient condition for satisfaction of the noncontextuality inequality is then

$$
1 - \Pr(\Lambda_{\text{ind}})(1 - \beta(\Gamma_G,q)) \leq 1 - p_0 (1 - \beta(\Gamma_G,q)) \frac{R - \alpha(G,w)}{\alpha^*(G,w) - \alpha(G,w)},
$$

(72)

which becomes

$$
p_0 \leq \frac{\alpha^*(G,w) - \alpha(G,w)}{R - \alpha(G,w)}\Pr(\Lambda_{\text{ind}})
$$

(73)

when $R > \alpha(G,w)$. Noting that

$$
R \leq \alpha(G,w) + \Pr(\Lambda_{\text{ind}})(\alpha^*(G,w) - \alpha(G,w)),
$$

we have

$$
\Pr(\Lambda_{\text{ind}}) \geq \frac{R - \alpha(G,w)}{\alpha^*(G,w) - \alpha(G,w)},
$$

(74)

so that the sufficient condition for satisfaction of the noncontextuality inequality becomes $p_0 \leq 1$, which is trivially satisfied.

When $R = \alpha(G,w)$, the sufficient condition becomes $\beta(\Gamma_G,q) \leq 1$, which is again trivially satisfied.

Finally, when $R < \alpha(G,w)$, the sufficient condition becomes

$$
p_0 \geq \frac{-\alpha^*(G,w) - \alpha(G,w)}{\alpha(G,w) - R}\Pr(\Lambda_{\text{ind}}),
$$

(75)

which is again trivially satisfied since $p_0 \geq 0$.

Hence trivial POVMs cannot yield a violation of our noncontextuality inequalities. This is the sense in which trivial POVMs cannot lead to nonclassicality in our approach, unlike the case of traditional Kochen-Specker approaches [9–11, 13]. To violate our noncontextuality inequalities, the POVMs must necessarily have some nontrivial projective component (that is not the identity operator or zero) but they need not be projectors.

### D. Open questions

We collect here the open questions raised in this paper, and also raise other open questions that merit further research:

1. **Characterizing structural Specker’s principle from probabilistic models on a hypergraph $\Gamma$:**

   Given that $\mathcal{C}^1(\Gamma) = \mathcal{G}(\Gamma)$ for some $\Gamma$, is it the case that $\Gamma$ must then necessarily satisfy structural Specker’s principle, namely, that every clique in $O(\Gamma)$ is a subset of some hyperedge in $\Gamma$? Or is it the case that there exists a hypergraph $\Gamma'$ for which $\mathcal{C}^1(\Gamma') = \mathcal{G}(\Gamma')$ but structural Specker’s principle fails?

   More generally, is there any characterization of a hypergraph satisfying structural Specker’s principle entirely in terms of the probabilistic models on it?

   As already pointed out earlier, this open question relates to the open Problem 7.2.3 of Ref. [11] of characterizing $\Gamma$ for which $\mathcal{C}^1(\Gamma) = \mathcal{G}(\Gamma)$. It is known that $\Gamma$ representing bipartite Bell scenarios [47] satisfy the property $\mathcal{C}^1(\Gamma) = \mathcal{G}(\Gamma)$ and we
have provided a generic recipe for converting any $\Gamma$ that does not satisfy structural Specker’s principle to a $\Gamma'$ that does satisfy it so that $C\mathcal{E}_1^{\mathcal{G}}(\Gamma') = C\mathcal{G}(\Gamma')$. The question is if there are any other $\Gamma$ that also satisfy $C\mathcal{E}_1^{\mathcal{G}}(\Gamma) = C\mathcal{G}(\Gamma)$.

2. Conditions for saturating the noise-robust noncontextuality inequalities:

Does there exist at least one extremal probabilistic model on $\Gamma_G$ (corresponding to an indeterministic vertex of the polytope, $\lambda \in \Lambda_{\text{ind}}$) such that $R(\lambda) = \alpha^*(G, w)$ and $\text{Corr}(\lambda) = \beta(\Gamma_G, q)$, for all $\Gamma_G$ of interest, i.e., $\Gamma_G$ such that $C(\Gamma_G) \neq \emptyset$ and $C\mathcal{E}_1^{\mathcal{G}}(\Gamma_G) = C\mathcal{G}(\Gamma_G)$?

If it doesn’t hold for some $\Gamma_G$, is it still possible: 1) to obtain a different noncontextual ontological model that saturates Eq. (32), or 2) to derive a (tight) noncontextuality inequality that is possibly different from Eq. (32) but is saturated by a non-contextual ontological model?

3. Properties of the weighted max-predictability, $\beta(\Gamma_G, q)$:

Since the crucial new hypergraph-theoretic ingredient in our inequalities is the weighted max-predictability, it would be interesting to understand properties of this hypergraph invariant on both counts: as a new mathematical object in its own right, one we haven’t been able to find a reference to in the hypergraph theory literature, as well as an important parameter of a hypergraph relevant for noise-robustness of a noise-robust noncontextuality inequality. Indeed, as we point out in footnote 23, identifying a distribution $q$ (in the definition of $\text{Corr}$, Eq. (19)) that minimizes $\beta(\Gamma_G, q)$ for a given $\Gamma_G$ would lead to better noise-robustness in the inequalities of Eqs. (32) or (33).

4. Noise-robust applications of quantum protocols based on KS-contextuality:

A general research direction is to construct noise-robust versions of applications that have previously been suggested for KS-contextuality. Our approach provides a recipe for doing this for any Bell-KS inequality appearing in such applications. Besides serving as a witness for strong nonclassicality [18] (i.e., Spekkens contextuality) noise-robust versions of these applications can help benchmark the experiments in terms of the noise that can be tolerated while still witnessing nonclassicality. Examples of such applications include those from Refs. [50–55].

\[ \text{A product experimental scenario, } \]  \[ \text{A general research direction is to construct noise-robust versions of applications that have previously been suggested for KS-contextuality. Our approach provides a recipe for doing this for any Bell-KS inequality appearing in such applications. Besides serving as a witness for strong nonclassicality [18] (i.e., Spekkens contextuality) noise-robust versions of these applications can help benchmark the experiments in terms of the noise that can be tolerated while still witnessing nonclassicality. Examples of such applications include those from Refs. [50–55].} \]

\[ \text{Appendix A: Trivial POVMs} \]

1. Bell-CHSH scenario

We have the Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$ for Alice ($\mathcal{H}_A$) and Bob ($\mathcal{H}_B$). Consider four binary-outcome POVMs, $\{A^{(0)}, A^{(1)}, B^{(0)}, B^{(1)}\}$, where

\[ A^{(0)} = \{A^{(0)}_0, A^{(0)}_1\}, \]
\[ A^{(1)} = \{A^{(1)}_0, A^{(1)}_1\}, \]
\[ B^{(0)} = \{B^{(0)}_0, B^{(0)}_1\}, \]
\[ B^{(1)} = \{B^{(1)}_0, B^{(1)}_1\}. \]  (A1)

0 ≤ $A^{(0)}_0$, $A^{(1)}_0$ ≤ $\mathbb{I}_{\mathcal{H}_A}$, 0 ≤ $B^{(0)}_0$, $B^{(0)}_1$ ≤ $\mathbb{I}_{\mathcal{H}_B}$, $A^{(0)}_0 + A^{(1)}_0 = A^{(1)}_0 + A^{(1)}_1 = \mathbb{I}_{\mathcal{H}_A}$, and $B^{(0)}_0 + B^{(1)}_0 = B^{(1)}_0 + B^{(1)}_1 = \mathbb{I}_{\mathcal{H}_B}$.

The quantum probability, given a shared quantum state $\rho_{AB}$ defined on $\mathcal{H}_A \otimes \mathcal{H}_B$, is given by

\[ p(a, b|x, y) = \text{Tr}(\rho_{AB} A^{(x)}_a \otimes B^{(y)}_b), \]  (A2)

\[ \text{Acknowledgments} \]

I would like to thank Andreas Winter for his comments on an earlier version of some of these ideas, Tobias Fritz for the ping-pong and the sing-song in which we often talked about hypergraphs, Rob Spekkens for the often argumentative – but always productive – conversations over lunch, and participants at the Contextuality conference (COCIUSA) at Perimeter Institute, during July 24 - 28, 2017, for very stimulating discussions that fed into the narrative of this paper. I would also like to thank David Schmid, Ana Belén Sainz, Elie Wolfe, and Tomáš Gonda for helping me better articulate the difference between structural vs. statistical readings of Specker’s principle, and Eric Cavalcanti for comments on the manuscript. Research at Perimeter Institute is supported by the Government of Canada through the Department of Innovation, Science and Economic Development Canada, and by the Province of Ontario through the Ministry of Research, Innovation and Science.

\[ \text{Appendix A: Trivial POVMs} \]

1. Bell-CHSH scenario

We have the Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$ for Alice ($\mathcal{H}_A$) and Bob ($\mathcal{H}_B$). Consider four binary-outcome POVMs, $\{A^{(0)}, A^{(1)}, B^{(0)}, B^{(1)}\}$, where

\[ A^{(0)} = \{A^{(0)}_0, A^{(0)}_1\}, \]
\[ A^{(1)} = \{A^{(1)}_0, A^{(1)}_1\}, \]
\[ B^{(0)} = \{B^{(0)}_0, B^{(0)}_1\}, \]
\[ B^{(1)} = \{B^{(1)}_0, B^{(1)}_1\}. \]  (A1)

0 ≤ $A^{(0)}_0$, $A^{(1)}_0$ ≤ $\mathbb{I}_{\mathcal{H}_A}$, 0 ≤ $B^{(0)}_0$, $B^{(0)}_1$ ≤ $\mathbb{I}_{\mathcal{H}_B}$, $A^{(0)}_0 + A^{(1)}_0 = A^{(1)}_0 + A^{(1)}_1 = \mathbb{I}_{\mathcal{H}_A}$, and $B^{(0)}_0 + B^{(1)}_0 = B^{(1)}_0 + B^{(1)}_1 = \mathbb{I}_{\mathcal{H}_B}$.

The quantum probability, given a shared quantum state $\rho_{AB}$ defined on $\mathcal{H}_A \otimes \mathcal{H}_B$, is given by

\[ p(a, b|x, y) = \text{Tr}(\rho_{AB} A^{(x)}_a \otimes B^{(y)}_b), \]  (A2)
for $a, b, x, y \in \{0, 1\}$. Here $A(x) \otimes B(y)$ is jointly measurable with $\mathbb{I}_{\mathcal{G}} \otimes B(y)$, just because of the commutativity of their respective POVM elements. The joint observable being measured is $A(x) \otimes B(y)$. Now, consider the case when all the POVM elements are trivial, i.e., $A_a = q_a^T \mathbb{I}_A$ and $B_b = \rho_b^T \mathbb{I}_B$, for some $q_a^T, \rho_b^T \in [0, 1]$ for all $a, b, x, y \in \{0, 1\}$. We then have

\[ p(a, b|x, y) = q_a^T \rho_b^T, \forall a, b, x, y \in \{0, 1\}. \quad (A3) \]

A global joint probability distribution which reproduces the above as marginals is simply given by their product:

\[ p(a, b|x, y) = q_a^T(q_b^T(0) r_b^T(0) r_b^T(1)). \quad (A4) \]

Hence, trivial POVMs never violate any Bell-CHSH inequality for this scenario.

2. CHSH-type contextuality scenario: 4-cycle

We now consider the Bell-CHSH scenario without the constraint of spacelike separation. What the lack of spacelike separation means from the quantum perspective is that one no longer needs to model this spacelike separation by requiring a tensor product structure, or (more generally) by requiring the commutativity of the observables that are jointly measured [13]. That is, there is no physical justification for imposing the tensor product structure or the commutativity of jointly measured observables.\(^{25}\)

Thus, we have the Hilbert space $\mathcal{H}$ and we consider four binary-outcome POVMs, $\{A(0), A(1), B(0), B(1)\}$, on $\mathcal{H}$, where

\[ A(0) \equiv \{0, 0\}, A(1) \equiv \{0, 1\}, B(0) \equiv \{0, 0\}, B(1) \equiv \{0, 1\}, \]

\[ 0 \leq A(0), A(1), B(0), B(1) \leq \mathbb{I}_\mathcal{H}, A(0) + A(1) = A(0) + A(1) = B(0) + B(1) = B(0) + B(1) = \mathbb{I}_\mathcal{H}. \]

Further, the following sets of POVMs are jointly measurable: $\{A(0), B(0)\}$, $\{A(0), B(1)\}$, $\{A(1), B(0)\}$, $\{A(1), B(1)\}$. The most general joint observable for a pair of compatible POVMs $\{A(x), B(y)\}$ is given by a POVM $G^{(xy)} \equiv \{G_{00}^{(xy)}, G_{01}^{(xy)}, G_{10}^{(xy)}, G_{11}^{(xy)}\}$ (that isn’t necessarily unique [17]) such that: $G_{00}^{(xy)} + G_{01}^{(xy)} = A(x) + G_{10}^{(xy)} + G_{11}^{(xy)} = B(y)$, $G_{00}^{(xy)} + G_{01}^{(xy)} = B(y)$, $G_{01}^{(xy)} + G_{11}^{(xy)} = B(y)$. In particular, if (and only if) the POVMs $A(x)$ and $B(y)$ commute, we can construct the joint POVM as a product: $G_{ab}^{(xy)} = A(x)B(y)$ for all $a, b, x, y \in \{0, 1\}$. In the absence of such commutativity, the joint POVM cannot be written as a product.

The quantum probability, given a quantum state $\rho$ on $\mathcal{H}$, is given by

\[ p(a, b|x, y) = \text{Tr}(\rho G_{ab}^{(xy)}), \quad (A6) \]

for $a, b, x, y \in \{0, 1\}$. Note that this probability depends on the joint measurement $G^{(xy)}$ implementing $A(x)$ and $B(y)$ together, and that, in general, there may be multiple choices of $G^{(xy)}$ possible. This is easy to see since there is one undetermined positive operator in the joint measurement that is not fixed by $A(x)$ or $B(y)$, i.e., we can write the POVM elements of $G^{(xy)}$ as: $G_{00}^{(xy)} = A(x) - G_{00}^{(xy)}$, $G_{10}^{(xy)} = B(y) - G_{00}^{(xy)}$, $G_{11}^{(xy)} = I - A(x) - B(y) + G_{00}^{(xy)}$, where $G_{00}^{(xy)}$ is a positive semidefinite operator satisfying $A(x) + B(y) - \mathbb{I}_\mathcal{H} \leq G_{00}^{(xy)} \leq A(x), B(y)$. Here $G_{00}^{(xy)}$ represents the freedom in the choice of how the joint measurement might be implemented within quantum theory. This freedom reflects the fact that since the jointly measured observables are no longer spacelike separated, it is possible to introduce correlations between them that are stronger than what is allowed in the corresponding Bell scenario in quantum theory. The strength of these correlations is only limited by the constraints on $G_{00}^{(xy)}$ imposed by the marginal observables $A(x)$ and $B(y)$. This is in contrast to the case where $A(x)$ and $B(y)$ are space-like separable observables and the only choice of joint POVM consistent with spacelike separation is fixed by $G_{00}^{(xy)} = A(x)B(y)$, i.e., the strength of correlations between $A(x)$ and $B(y)$ is fixed entirely by them and there is no freedom in choosing $G^{(xy)}$.

Thus, we have that $A(x)$ is jointly measurable with $B(y)$ and $G^{(xy)}$ denotes a joint POVM of $A(x)$ and $B(y)$. Now, consider the case when all the POVM elements are trivial, i.e., $A_a = q_a^T \mathbb{I}_\mathcal{H}$ and $B_b = \rho_b^T \mathbb{I}_\mathcal{H}$, for some $q_a^T, \rho_b^T \in [0, 1]$ for all $a, b, x, y \in \{0, 1\}$. A possible joint POVM for these trivial POVMs is then the product POVM:

\[ G_{ab}^{(xy)} = A_a B_b = \frac{1}{4} \mathbb{I}_\mathcal{H}. \quad (A7) \]

If one restricted joint measurability of $A(x)$ and $B(y)$ to just commutativity — a sufficient but not necessary

\(^{25}\)On the other hand, what this lack of spacelike separation means from the perspective of an ontological model is that one no longer has a justification for assuming factorizability [17] and, consequently, the generalization of Fine’s theorem [35] fails to prove that there is no loss of generality in assuming outcome determinism in discussions of KS-contextuality (unlike the case of Bell scenarios, where factorizability is justified by spacelike separation); there is a definite loss of generality, in that measurement noncontextual and outcome-ndeterministic ontological models that are non-factorizable are not empirically equivalent to measurement noncontextual and outcome-deterministic (or KS-noncontextual) ontological models. See Ref. [32] for a discussion of this aspect.
condition for joint measurability[26] — we would take the above choice of the product POVM as a “natural” one. Being a product of trivial POVMs, this choice will never lead to a violation of the CHSH-type inequality for this scenario. Indeed, the structure of a Bell scenario — requiring the decomposition of the Hilbert space as $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ (tensor product paradigm), or more generally, imposing the commutativity requirement $[A_a^{(x)} B_b^{(y)}] = 0$ (commutativity paradigm) — is such that the only possible choice of joint measurement that can be implemented by spacelike separated parties is the one that corresponds to the product POVM, given by operators $G_{ab}^{(xy)} = A_a^{(x)} B_b^{(y)}$.

However, this is not the only allowed joint measurement for these trivial POVMs, particularly when there is no locality constraint on the measurements from spacelike separation[27]. An extreme choice of joint POVM is the following:

$$G_{ab}^{PR(xy)} = \frac{I_y}{2} \delta_{a \oplus b, xy}, \quad (A8)$$

which leads to the probability distribution $p(a, b|x, y) = \frac{1}{2} \delta_{a \oplus b, xy}$ for any choice of quantum state. Hence, this joint POVM $G_{ab}^{PR(xy)}$ always yields statistics corresponding to the PR-box, maximally violating the CHSH-type inequality for this scenario, namely,

$$\sum_{a, b, x, y | a \oplus b = xy} \frac{1}{4} p(a, b|x, y) \leq \frac{3}{4}. \quad (A9)$$

Physically, it’s possible to implement this (without requiring any quantum resources) by providing a box that always produces these correlations between measurement settings denoted by $(xy) \in \{0, 1\}^2$, regardless of the input state. Such a black-box would maximally violate the CHSH-type inequality (viewed as a Bell-KS inequality witnessing KS-contextuality), but that shouldn’t be surprising in the absence of spacelike separation. Also, the trivial PR-box joint POVM $G_{ab}^{PR(xy)}$ is a perfectly valid way to implement the joint measurement of trivial POVMs $A^{(x)}$ and $B^{(y)}$ within the standard paradigm of operational quantum theory[28].

To summarize, we note the following:

- Within the traditional framework of KS-noncontextuality, if one wants to go beyond projective measurements to arbitrary POVMs in a contextuality scenario, then one must — in order to avoid the pathology of trivial POVMs violating the Bell-KS inequalities maximally — restrict by fiat the notion of joint measurability to merely commutativity. This is, for example, the attitude adopted in Ref. [13].

- However, if one is going beyond projective measurements, we know that commutativity is only a sufficient condition for joint measurability, not a necessary one[29].

- This brings us to our observation that the traditional notion of KS-noncontextuality is pathologically once the most general situation in quantum theory is considered: arbitrary POVMs with the general notion of joint measurability (see, e.g., Ref. [29] for this notion and its relation to commutativity). In particular, in the absence of spacelike separation, there is no physical justification to restrict the notion of joint measurability to merely commutativity.

- A similar consideration applies at the level of a KS-noncontextual ontological model: there, factorizability is not justified in the absence of spacelike separation. So, on those grounds alone, one should go beyond KS-noncontextuality as one’s notion of classicality; particularly, if one wants a notion of classicality that does not presume outcome determinism, just as local causality doesn’t presume it. This was argued in Ref. [30]. We point out the pathology of trivial POVMs only to drive home this point in a different way.

Appendix B: The KS-uncolourable hypergraph $\Gamma_{18}$

It is instructive to consider the KS-uncolourable hypergraph $\Gamma_{18}$, originally appearing in Ref. [33], and studied

[26] Particularly in the absence of spacelike separation. It is the need to model spacelike separation in a quantum Bell experiment that makes commutativity a necessary (and sufficient) condition for joint measurability of spacelike separated observables in a Bell scenario.

[27] To incorporate such a constraint, spacelike separation needs to be modelled via either the tensor product paradigm or the commutativity paradigm. Both these ways of modelling spacelike separation lead to the same set of quantum correlations for any finite-dimensional Hilbert space $\mathcal{H}$. The question of whether the two paradigms lead to the same set of correlations in the case of infinite dimensional Hilbert spaces is the subject of Tsirelson’s problem [56, 57]. Most studies of Bell-nonlocality are primarily concerned with finite dimensional Hilbert spaces; should one encounter infinite dimensional Hilbert spaces, the commutativity paradigm is the proper way to model spacelike separation.

[28] Note that the point of this demonstration is to show how, in the absence of spacelike separation justifying commutativity or a promise that the measurements are sharp, arbitrary correlations are achievable in quantum theory if unsharp measurements are allowed. All trivial POVMs are unsharp, but the converse is not true. That is, one can consider nontrivial POVMs that don’t violate the CHSH-type inequality maximally, but which violate it (arbitrarily) more than is allowed by sharp measurements in quantum theory. One could construct them, for example, by just taking a convex combination of the PR-box trivial POVM with some sharp (and thus product) joint POVM.
in the light of Spekkens contextuality in Ref. [1]. This hypergraph fails both criteria for the hypergraphs $\Gamma$ considered in this paper, namely, $\mathcal{C}(\Gamma) \neq \emptyset$ (KS-colourability) and $\mathcal{C}\mathcal{E}^1(\Gamma) = \mathcal{G}(\Gamma)$.

For probabilistic models on $\Gamma_{18}$, the following hold: $\mathcal{C}(\Gamma_{18}) = \emptyset \subsetneq \mathcal{C}\mathcal{E}^1(\Gamma_{18}) \subsetneq \mathcal{G}(\Gamma_{18})$. This was considered in Ref. [1], where $\mathcal{C}\mathcal{E}^1(\Gamma_{18})$ excludes the extremal probabilistic model in $\mathcal{G}(\Gamma_{18})$ that corresponds to the upper bound on the noise-robust noncontextuality inequality of Ref. [1]. As argued in Ref. [1], this noise-robust noncontextuality inequality is the appropriate operational generalization (to possibly noisy measurements) of the Kochen-Specker contradiction first demonstrated in Ref. [13]; this generalization cannot be accommodated in our generalization of the CSW framework [14].

If one extends the KS-uncolourable $\Gamma_{18}$ to a KS-colourable hypergraph $\Gamma_{27}$ with 9 “no-detection” events, one for each hyperedge, then we have $\mathcal{C}(\Gamma_{27}) \neq \emptyset$, but it’s still the case that $\mathcal{C}(\Gamma_{27}) \subsetneq \mathcal{C}\mathcal{E}^1(\Gamma_{27}) \subsetneq \mathcal{G}(\Gamma_{27})$ for this hypergraph. $\Gamma_{27}$ Hence, $\Gamma_{27}$ cannot be understood in our generalization of the CSW framework either.

Indeed, if one “blindly” writes down a CSW classical bound for some Bell-KS expression defined on $O(\Gamma_{18})$, then such a bound is equivalently a bound for the same Bell-KS expression defined on $\Gamma_{27}$ (where normalization is restored). Further, the E1 bound on $\Gamma_{18}$ is a CE$^1$ bound on $\Gamma_{27}$. The GPT bound happens to agree with the CE$^1$ bound for a particular Bell-KS expression (sum of all probabilities) but differs for some other Bell-KS expressions defined on this hypergraph. Consider, for example, the following three expressions (see Fig. 8):

\[
\text{Expr}_1 \equiv \sum_{v \in V(\Gamma_{18})} p(v), \\
\text{Expr}_2 \equiv \sum_{v \in V(\Gamma_3)} p(v), \\
\text{Expr}_3 \equiv \sum_{v \in V(\Gamma_{18})} p(v) + \sum_{v \in V(\Gamma_3)} p(v). \quad (B1)
\]

We have:

\[
\begin{align*}
\text{Expr}_1 & \leq 8 < 9 = \text{Expr}_2, \\
\text{Expr}_2 & \leq 1 < \frac{3}{2}, \\
\text{Expr}_3 & \leq 9 < 10 < 10.5. \quad (B2)
\end{align*}
\]

29 This follows from noting that extremal probabilistic models on $\Gamma_{18}$ are still extremal probabilistic models on $\Gamma_{27}$: ones where the no-detection events are assigned zero probabilities. See Theorem 2.5.3 of Ref. [11].

30 Note that adding these no-detection events is equivalent to allowing subnormalized probabilities (i.e., sum of probabilities assigned to measurement events in a hyperedge can be less than 1) on $\Gamma_{18}$. Hence, even allowing for subnormalization on $\Gamma_{18}$, which means that one is looking at probabilistic models on the hypergraph $\Gamma_{27}$, does not eliminate the gap between CE$^1$ probabilistic models and general probabilistic models, so that any upper bound on a Bell-KS expression given by probabilistic models in $\mathcal{C}\mathcal{E}^1(\Gamma_{27})$ is not always the same as the general probabilistic upper bound from probabilistic models in $\mathcal{G}(\Gamma_{27})$. The CSW framework only considers the upper bound given by $\mathcal{C}\mathcal{E}^1(\Gamma_{27})$ probabilistic models.
Thus, Expr_3 is a Bell-KS expression that discriminates between probabilistic models at all three levels of the hierarchy. Indeed, the upper bound on Expr_3 for CE^3(Γ_{27}) models can be saturated by projective quantum realizations of the hypergraph, in particular the standard realization with 18 rays, with the zero operator for the no-detection events [43]. The fact that there exists such a Bell-KS expression as Expr_3 means that the CE^3 upper bounds from the CSW approach can be violated by a general probabilistic model, i.e., the upper bounds for CE^3 models and general probabilistic models don’t agree, and we cannot take the graph-theoretic upper bounds of CSW for granted in our noise-robust noncontextuality inequalities. Indeed, the general probabilistic upper bound for any Bell-KS expression defined on a contextuality scenario is a hypergraph invariant — in the sense that it is a property that is shared by all hypergraphs isomorphic to each other — that may or may not be expressible as a graph invariant à la CSW.

What, then, do the bounds given by graph invariants of CSW for O(Γ_{18}) mean in our generalization of the CSW framework? Following our approach, outlined in Sec. III.B, we can go from G = O(Γ_{18}) to the hypergraph Γ_G = Γ_{O(Γ_{18})} (see Fig. 9) for which we have (by construction) C(Γ_{O(Γ_{18})}) ≠ ∅ (so that the underlying hypergraph is no longer KS-uncolourable) and CE^3(Γ_{O(Γ_{18})}) = G(Γ_{O(Γ_{18})}) (so that, for any Bell-KS expression, the upper bound given by the fractional packing number α*(G, w) in the CSW framework agrees with the general probabilistic upper bound). Since this construction proceeds by converting all maximal cliques in Γ_{18} to hyperedges in Γ_{O(Γ_{18})} and adding a new vertex to each such hyperedge, it achieves both purposes: firstly, adding a (no-detection) vertex to every maximal clique that is a hyperedge in Γ_{18} ensures the KS-colourability of Γ_{O(Γ_{18})}, i.e., C(Γ_{O(Γ_{18})}) ≠ ∅, and secondly, adding a vertex to every maximal clique that is not a hyperedge in Γ_{18} ensures that CE^3(Γ_{O(Γ_{18})}) = G(Γ_{O(Γ_{18})}). Once these two properties are satisfied, the graph invariants of CSW [10] become applicable to any Bell-KS expression defined for any set of vertices in the subhypergraph Γ_{18} of Γ_{O(Γ_{18})}.

Our noise-robust noncontextuality inequality then applies to the KS-colourable hypergraph Γ_{O(Γ_{18})}, where the graph invariants of CSW make sense, rather than the KS-uncolourable hypergraph Γ_{18}. On the other hand, an appropriate noise-robust noncontextuality inequality for the KS-uncolourable hypergraph Γ_{18} is, then, the one reported in Ref. [11].

[1] R. Kunjwal and R. W. Spekkens, “From the Kochen-Specker theorem to noncontextuality inequalities without assuming determinism”, Phys. Rev. Lett. 115, 110403 (2015).
[2] M. D. Mazurek, M. F. Pusey, R. Kunjwal, K. J. Resch, R. W. Spekkens, “An experimental test of noncontextuality without unphysical idealizations”, Nat. Commun. 7, 11780 (2016).
[3] A. Krishna, R. W. Spekkens, and E. Wolfe, “Deifying robust noncontextuality inequalities from algebraic proofs of the Kochen-Specker theorem: the Peres-Mermin square”, arXiv:1704.01153 [quant-ph] (2017).
[4] D. Schmid and R. W. Spekkens, “Contextual advantage for state discrimination”, arXiv:1706.04588 [quant-ph] (2017).
[5] R. Kunjwal and R. W. Spekkens, “From statistical proofs of the Kochen-Specker theorem to noise-robust non-contextuality inequalities”, arXiv:1708.04793 [quant-ph] (2017).
[6] D. Schmid, R. W. Spekkens, and E. Wolfe, “All the noncontextuality inequalities for arbitrary prepare-and-measure experiments with respect to any fixed sets of operational equivalences”, arXiv:1710.08434 [quant-ph] (2017).
[7] R. W. Spekkens, “Contextuality for preparations, transformations, and unsharp measurements”, Phys. Rev. A 71, 052108 (2005).

31 The approach for KS-uncolourable hypergraphs will be further developed in hypergraph-theoretic terms in forthcoming work [23].

[8] S. Kochen and E. P. Specker, “The Problem of Hidden Variables in Quantum Mechanics”, J. Math. Mech. 17, 59 (1967).
[9] A. Cabello, S. Severini, and A. Winter, “(Non-)Contextuality of Physical Theories as an Axiom”, arXiv:1010.2163 [quant-ph] (2010).
[10] A. Cabello, S. Severini, and A. Winter, “Graph-Theoretic Approach to Quantum Correlations”, Phys. Rev. Lett. 112, 040401 (2014).
[11] A. Acín, T. Fritz, A. Leverrier, and A. B. Sainz, A Combinatorial Approach to Nonlocality and Contextuality, Comm. Math. Phys. 334(2), 533-628 (2015).
[12] J. Barrett, “Information processing in generalized probabilistic theories”, Phys. Rev. A 75, 032304 (2007).
[13] S. Abramsky and A. Brandenburger, “The Sheaf-Theoretic Structure Of Non-Locality and Contextuality”, New Journal of Physics 13 (2011) 113036.
[14] Y. C. Liang, R. W. Spekkens, H. M. Wiseman, “Specker’s parable of the overprotective seer: A road to contextuality, nonlocality and complementarity”, Phys. Rep. 506, 1 (2011).
[15] R. Kunjwal and S. Ghosh, “Minimal state-dependent proof of measurement contextuality for a qubit”, Phys. Rev. A 89, 042118 (2014).
[16] R. Kunjwal, C. Heunen, and T. Fritz, “Quantum realization of arbitrary joint measurability structures”, Phys. Rev. A 89, 052126 (2014).
[17] R. Kunjwal, “A note on the joint measurability of POVMs and its implications for contextuality”, arXiv:1403.0470 [quant-ph] (2014).
[18] J. S. Bell, “On the Einstein-Podolsky-Rosen Paradox”, Physics 1, 195 (1964). Reprinted in Ref. [55], chap. 2.
[19] J. S. Bell, “On the problem of hidden variables in quantum mechanics”, Rev. Mod. Phys. 38, 447 (1966); Reprinted in Ref. [58], chap. 1.
[20] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, “Proposed Experiment to Test Local Hidden-Variable Theories”, Phys. Rev. Lett. 23, 880 (1969).
[21] B. Hensen et al., “Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres”, Nature 526, 682686 (2015).
[22] Lynden K. Shalm et al., Strong Loophole-Free Test of Local Realism, Phys. Rev. Lett. 115, 250402 (2015).
[23] M. Giustina et al., “Significant-Loophole-Free Test of Bell Theorem with Entangled Photons”, Phys. Rev. Lett. 115, 250401 (2015).
[24] S. Popescu and D. Rohrlich, “Quantum nonlocality as an axiom”, Found Phys (1994) 24: 379.
[25] A. Fine, “Hidden Variables, Joint Probability, and the Bell Inequalities”, Phys. Rev. Lett. 48, 291 (1982).
[26] R. Chaves and T. Fritz, “Entropic approach to local realism and noncontextuality”, Phys. Rev. A 85, 032113 (2012).
[27] Tobias Fritz, Rafael Chaves, “Entropic Inequalities and Marginal Problems”, IEEE Trans. on Information Theory, vol. 59, pages 803 - 817 (2013).
[28] A. Cabello, “What do we learn about quantum theory from the Kochen-Specker quantum contextuality?”, PIRSA:17070034 (2017).
[29] T. Heinosaari, D. Reitzner, and P. Stano, “Notes on Joint Measurability of Quantum Observables”, Foundations of Physics (2008) 38: 1133-1147.
[30] R. Kunjwal, “Fine’s theorem, noncontextuality, and correlations in Specker’s scenario”, Phys. Rev. A 91, 022108 (2015).
[31] G. Chiribella and X. Yuan, “Measurement sharpness cuts nonlocality and contextuality in every physical theory”, arXiv:1404.3348 [quant-ph] (2014).
[32] G. Chiribella and X. Yuan, “Bridging the gap between general probabilistic theories and the device-independent framework for nonlocality and contextuality”, Information and Computation, 250, 15-49 (2016), arXiv:1504.02395 [quant-ph].
[33] R. W. Spekkens, “Noncontextuality: how should we define it, why it is natural, and what to do about its failure”, PIRSA:17070035, http://pirsa.org/17070035/.
[34] M. D. Mazurek, M. F. Pusey, K. J. Resch, and R. W. Spekkens, “Experimentally bounding deviations from quantum theory in the landscape of generalized probabilistic theories”, arXiv:1710.05948 [quant-ph] (2017).
[35] R. Kunjwal, “How to go from the KS theorem to experimentally testable noncontextual inequalities”, PIRSA:17070059, http://pirsa.org/17070059/.
[36] N. Harrigan and R. W. Spekkens, “Einstein, incompleteness, and the epistemic view of quantum states,” Found. Phys. 40, 125 (2010).
[37] Konrad Engel, “Sperner theory, Encyclopedia of Mathematics and its Applications”, vol. 65, Cambridge University Press, Cambridge, 1997.
[38] A. A. Klyachko, M. A. Can, S. Binicioğlu, and A. S. Shumovsky, Simple Test for Hidden Variables in Spin-1 Systems, Phys. Rev. Lett. 101, 020403 (2008).
[39] A. Cabello, “Specker’s fundamental principle of quantum mechanics”, arXiv:1212.1756 [quant-ph] (2012).
[40] T. Gonda, R. Kunjwal, D. Schmid, E. Wolfe, and A. B. Sainz, “Almost Quantum Correlations are Inconsistent with Specker’s Principle”, arXiv:1712.01225 [quant-ph] (2017).
[41] G. Chiribella, G. M. D’Ariano, and P. Perinotti, “Probabilistic theories with purification”, Phys. Rev. A 81, 062348 (2010).
[42] M. Navascués, Y. Guryanova, M. J. Hoban, and A. Acín, “Almost quantum correlations”, Nature Communications 6, 6288 (2015).
[43] A. Cabello, Adan, J. Estebaran, and G. Garcia-Alcaine, “Bell-Kochen-Specker theorem: A proof with 18 vectors,” Physics Letters A 212, 183 (1996).
[44] R. Kunjwal, “Irreducible noncontextuality inequalities from logical proofs of the Kochen-Specker theorem,” (forthcoming).
[45] X. Zhan, E. G. Cavalcanti, J. Li, Z. Bian, Y. Zhang, H. M. Wiseman, and P. Xue, “Experimental generalized contextuality with single-photon qubits”, Optica 4, 966-971 (2017).
[46] R. Kunjwal, “Contextuality beyond the Kochen-Specker theorem”, arXiv:1612.07250 [quant-ph] (2016).
[47] T. Fritz, A. B. Sainz, R. Augusiak, J. B. Brask, R. Chaves, A. Leverrier, and A. Acín, “Local orthogonality: a multiparticle principle for correlations”, Nat. Commun. 4, 2263 (2013).
[48] R. W. Spekkens, “Nonclassicality as the failure of noncontextuality”, http://pirsa.org/15050081/ (see the slide at 41:43 minutes).
[49] R. W. Spekkens, “Quasi-quantization: classical statistical theories with an epistemic restriction”, arXiv:1409.5041 [quant-ph] (2014).
[50] T. Vidick and S. Wehner, “Does Ignorance of the Whole Imply Ignorance of the Parts? Large Violations of Noncontextuality in Quantum Theory”, Phys. Rev. Lett. 107, 030402 (2011).
[51] R. Raussendorf, “Contextuality in measurement-based quantum computation”, Phys. Rev. A 88, 022322 (2013).
[52] M. Howard, J. Wallman, V. Veitch, and J. Emerson, “Contextuality supplies the magic for quantum computation”, Nature 510, 351 (2014).
[53] N. Delfosse, P. A. Guerin, J. Bian, and R. Raussendorf, “Wigner Function Negativity and Contextuality in Quantum Computation on Rebits”, Phys. Rev. X 5, 021003 (2015).
[54] J. Bermejo-Vega, N. Delfosse, D. E. Browne, C. Okay, R. Raussendorf, “Contextuality as a resource for qubit quantum computation”, arXiv:1610.08529 [quant-ph] (2016).
[55] J. Singh, K. Bharti, and Arvind, Quantum key distribution protocol based on contextuality monogamy, Phys. Rev. A 95, 062333 (2017).
[56] V. B. Scholz and R. F. Werner, “Tsirelson’s Problem”, arXiv:0812.4305 [math-ph] (2008).
[57] T. Fritz, “Tsirelson’s problem and Kirchberg’s conjecture”, Rev. Math. Phys. 24(5), 1250012 (2012).
[58] J. S. Bell, “Speakable and unspeakable in quantum mechanics” (Cambridge University Press, New York, 1987).