Fine's theorem, noncontextuality, and correlations in Specker's scenario

Ravi Kunjwal[*]

Optics & Quantum Information Group, The Institute of Mathematical Sciences, C.I.T Campus, Taramani, Chennai 600 113, India
(Received 28 October 2014; published 12 February 2015)

A characterization of noncontextual models which fall within the ambit of Fine's theorem is provided. In particular, the equivalence between the existence of three notions is made explicit: a joint probability distribution over the outcomes of all the measurements considered, a measurement-noncontextual and outcome-deterministic (or KS-noncontextual, where KS stands for Kochen-Specker) model for these measurements, and a measurement-noncontextual and factorizable model for them. A KS inequality, therefore, is implied by each of these three notions. Following this characterization of noncontextual models that fall within the ambit of Fine's theorem, nonfactorizable noncontextual models which lie outside the domain of Fine's theorem are considered. While outcome determinism for projective (sharp) measurements in quantum theory can be shown to follow from the assumption of preparation noncontextuality, such a justification is not available for nonprojective (unsharp) measurements which ought to admit outcome-indeterministic response functions. The Liang-Spekkens-Wiseman (LSW) inequality is cited as an example of a noncontextuality inequality that should hold in any noncontextual model of quantum theory without assuming factorizability. Three other noncontextuality inequalities, which turn out to be equivalent to the LSW inequality under relabelings of measurement outcomes, are derived for Specker's scenario. The polytope of correlations admissible in this scenario, given the no-disturbance condition, is characterized.

I. INTRODUCTION

In attempts to provide a more complete description of reality than operational quantum theory in terms of a noncontextual ontological model, it is almost always assumed that whatever the ontic state λ is, it must specify the outcomes of measurements exactly (an assumption called *outcome determinism*) and any operational unpredictability in the measurement outcomes is on account of coarse graining over these ontic states λ. This paper concerns itself with what can still be said about noncontextuality if outcome determinism is not assumed: the ontic state is not always required to fix the outcomes of measurements but only their probabilities. The physical motivation for this becomes clear once the following questions are asked:

(1) Do there exist noncontextual ontological models of quantum theory where the ontic state λ fixes the outcomes of measurements?

The Kochen-Specker theorem [1] rules out this possibility. Let us now remove the requirement of outcome determinism, namely, that λ fix the outcomes of measurements, and ask the question:

(2) Do there exist noncontextual ontological models of quantum theory where the ontic state λ fixes the *probabilities* of outcomes of measurements?

The Kochen-Specker theorem [1] is silent on this question since it presumes the ontic state λ must fix the outcomes of (projective) measurements. This question is most naturally addressed in the framework of generalized noncontextuality due to Spekkens [2]. This is the framework adopted in this paper.

It is well known that, in contrast to the Kochen-Specker theorem [1], Bell's theorem [3–5] does not require an assumption that the ontic state λ fixes the outcomes of the measurements. This becomes particularly clear in view of Fine's theorem [6,7] that, in a Bell scenario, a locally deterministic model [3] exists if and only if a locally causal (or "Bell-local") model [4,5] exists, and how this is equivalent to requiring the existence of a joint probability distribution over outcomes of all the measurements considered in a Bell scenario. Hence, even if the outcomes are only determined probabilistically by λ in the local hidden variable model, Bell's theorem holds. The key issue in Bell scenarios is factorizability: the conditional independence of the outcomes of spacelike separated measurements given the ontic state λ of the system,

$$\xi(X_1, \ldots, X_N | M_1, \ldots, M_N, \lambda)$$
$$= \xi(X_1 | M_1, \lambda)\xi(X_2 | M_2, \lambda) \ldots \xi(X_N | M_N, \lambda), \quad (1)$$

where X_i labels the outcome of measurement M_i performed by the ith party, $i \in \{1, \ldots, N\}$. All these response functions may be outcome indeterministic, i.e., $\xi \in [0,1]$. Indeed, factorizability is a necessary consequence of any set of assumptions that may be used to derive Bell's theorem [5].

Note also that factorizability is a weaker constraint than outcome determinism since the latter implies the former but the converse does not hold: that is, given that $\xi \in \{0,1\}$ for all the response functions above, it follows that $\xi(X_1, \ldots, X_N | M_1, \ldots, M_N, \lambda) = \delta_{X_1 \cdots X_N, X'_1 \cdots X'_N(\lambda)}$ and $\xi(X_i | M_i, \lambda) = \delta_{X_i, X'_i(\lambda)}$ for all $i \in \{1 \cdots N\}$, where $X'_i(\lambda)$ is the outcome assigned to measurement M_i by λ. Obviously, then, $\xi(X_1, \ldots, X_N | M_1, \ldots, M_N, \lambda) = \delta_{X_1 \cdots X_N, X'_1 \cdots X'_N(\lambda)} = \prod_{i=1}^{N} \delta_{X_i, X'_i(\lambda)} = \prod_{i=1}^{N} \xi(X_i | M_i, \lambda)$.

On the other hand, things are not as straightforward for contextuality [1,8,9]. Mathematically, both Bell-local models and KS-noncontextual models rely on the existence of a joint probability distribution over all measurement outcomes in a given scenario such that this distribution reproduces the

---
[*]rkunj@imsc.res.in

observed statistics as marginals. Given this correspondence between Bell's theorem and the KS theorem, one may ask whether the assumption of outcome determinism is really required in the KS theorem and whether the KS theorem excludes also all outcome-indeterministic noncontextual models on account of Fine's theorem. This paper answers this question in the negative.

The outcome-indeterministic noncontextual models excluded by the KS theorem are precisely the ones where factorizability holds. However, in the absence of spacelike separation between measurements one does not have a compelling justification to assume that measurement outcomes are conditionally independent of each other given the ontic state λ. The physical meaning of factorizability is this: that the measurement outcomes do not have any correlations that are not due to the ontic state λ of the system. One could, on the other hand, imagine an adversarial situation where two measurements are correlated—which is physically possible if they are not spacelike separated—and this correlation is not mediated only by the ontic state λ of the system but is perhaps encoded in the degrees of freedom of the measurement apparatus by an adversary who wants to convince the experimenter that something nonclassical is going on (in the sense of KS contextuality) but, really, it is correlated noise that is doing all the work of violating a KS inequality. The LSW inequality [10,11] is an example of a noncontextuality inequality that takes this possibility into account and raises the bar for what correlations count as nonclassical. This is why we need to consider noncontextual models which are not factorizable. Since all KS-noncontextual models are factorizable on account of Fine's theorem, as will become clear in Sec. III, noncontextual models which are not factorizable are exclusively taken into account *only* in the generalized definition of noncontextuality [2]. This realization is a key conceptual insight of this paper, pointing to the necessity of revising the traditional analyses of KS noncontextuality to accommodate the generalized notion of noncontextuality [2].

Besides, just as local causality does not require the assumption of outcome determinism, a good definition of noncontextuality should also not appeal to outcome determinism (or even factorizability). Experimental violations of Bell inequalities certify a kind of nonclassicality independent of the truth of quantum theory, a feature that makes Bell inequality violations an invaluable resource in device-independent protocols [12]. In contrast, a KS-noncontextual model has to refer to projective (sharp) measurements in quantum theory and assume outcome determinism for them in order to obtain a KS inequality: neither of these is needed in a Bell-local model. The generalized notion of noncontextuality offers the possibility of taking about noncontextuality without making the assumption that the operational theory is quantum theory. The present paper, however, restricts itself to generalized noncontextuality for operational quantum theory.

The main contributions of this paper are twofold: First, as noted above and proven in Sec. III, it shows the relevance of Fine's theorem in recognizing the limitations that the assumption of outcome determinism places on considerations of noncontextual models of quantum theory, as well as why one should worry about nonfactorizable noncontextual models. Second, after noting these connections between Fine's theorem, noncontextuality, and the status of outcome determinism in noncontextual models of quantum theory [13], a complete analysis of Specker's scenario is provided in Sec. IV for nonfactorizable noncontextual models.

More specifically: In Sec. II, the notions of an operational theory and an ontological model of an operational theory are recalled, followed by the definition of noncontextuality due to Spekkens. Unless otherwise specified, "noncontextuality" will refer to this generalized notion in this paper. In particular, the traditional notion of noncontextuality, due to Bell, Kochen, and Specker [1,8,9], will be referred to as "KS noncontextuality." In Sec. III, I state Fine's theorem in the language of (generalized) noncontextuality [2] and discuss how this fits with the interpretation of Fine's theorem applied to Bell scenarios. In particular, this shows why outcome determinism is not an issue in Bell scenarios but it is an issue that needs to be handled with care in noncontextual models (see Ref. [13] for other compelling reasons for this). Section IV characterizes the correlations in Specker's scenario. I derive three noncontextuality inequalities that do not assume outcome determinism or factorizability. They turn out to be equivalent to the LSW inequality under appropriate relabeling of measurement outcomes in this scenario. Also, the polytope of correlations admissible in Specker's scenario, given the no-disturbance condition, is characterized by specifying all its extremal points. Section V concludes with a discussion.

II. OPERATIONAL THEORIES AND ONTOLOGICAL MODELS

Operational theory: An operational theory is specified by a triple $(\mathcal{P}, \mathcal{M}, p)$, where $\mathcal{P}$ denotes the preparation procedures $P \in \mathcal{P}$ in the laboratory, $\mathcal{M}$ denotes the measurement procedures $(M, \mathcal{K}_M) \in \mathcal{M}$, and $p : (\mathcal{P}, \mathcal{M}) \to [0,1]$ is the probability $p(k|P,M)$ that measurement outcome $k \in \mathcal{K}_M$ is observed when measurement procedure $(M, \mathcal{K}_M)$ is implemented following the preparation procedure P.

Ontological model: An ontological model (Λ, Ξ, μ) of an operational theory $(\mathcal{P}, \mathcal{M}, p)$ posits a space of ontic states $\lambda \in \Lambda$, probability densities $\mu_P : \Lambda \to [0,\infty)$ corresponding to preparation procedures $P \in \mathcal{P}$, and response functions $\xi : (\Lambda, \mathcal{M}) \to [0,1]$ denoting the probability $\xi(k|M,\lambda)$ that measurement outcome $k \in \mathcal{K}_M$ is observed when measurement procedure $(M, \mathcal{K}_M)$ is implemented and the ontic state of the system is λ. Note that $\int d\lambda \mu_P(\lambda) = 1$ and $\sum_{k \in \mathcal{K}_M} \xi(k|M,\lambda) = 1$. Ξ denotes the set of response functions in the ontological model for the measurement procedures in the operational theory. The ontological model must be empirically adequate, that is

$$p(k|P,M) = \int d\lambda \mu_P(\lambda) \xi(k|M,\lambda), \qquad (2)$$

for all $P \in \mathcal{P}, (M, \mathcal{K}_M) \in \mathcal{M}$. This is how an operational theory and its ontological model fit together.

Noncontextuality: An ontological model of an operational theory is defined to be noncontextual for prepare-and-measure experiments if it satisfies two properties: *preparation noncontextuality* and *measurement noncontextuality*. The content of preparation noncontextuality is captured in the following

inference from the operational theory to its ontological model:

$$p(k|P,M) = p(k|P',M), \forall k \in \mathcal{K}_M, \forall (M, \mathcal{K}_M) \in \mathcal{M}$$

$$\Rightarrow \mu_P(\lambda) = \mu_{P'}(\lambda), \forall \lambda \in \Lambda. \tag{3}$$

That is, two preparations P and P' which are operationally indistinguishable are represented by identical distributions in the ontological model. Similarly, measurement noncontextuality is simply expressed as the following inference:

$$p(k|P,M) = p(k|P,M'),$$

$$\forall k \in \mathcal{K}, (M, \mathcal{K}) \text{ and } (M', \mathcal{K}) \in \mathcal{M}, \forall P \in \mathcal{P} \tag{4}$$

$$\Rightarrow \xi(k|M,\lambda) = \xi(k|M',\lambda), \forall \lambda \in \Lambda. \tag{5}$$

That is, measurements M and M' which do not differ in their statistics relative to all preparations $P \in \mathcal{P}$ are represented by identical response functions in the ontological model. Fine's theorem, as formulated in this paper, relates to the assumption of measurement noncontextuality.

This assumption of noncontextuality is motivated by a methodological principle: do not introduce any differences in your explanation of two phenomena if no experiment can tell the phenomena apart. That is, if two experimental procedures are operationally indistinguishable then they should also be indistinguishable at the ontological level, also known as the (ontological) identity of (operational) indiscernibles. For further reading, I refer the reader to Ref. [2] where this notion of noncontextuality was first defined (see also, [13]) and its connection with the traditional notion of KS noncontextuality was also demonstrated.

Outcome determinism is the assumption that every response function in the ontological model is deterministic, i.e., $\xi(k|M,\lambda) \in \{0,1\}$ for all $(M, \mathcal{K}_M) \in \mathcal{M}$, $\lambda \in \Lambda$, and $\xi \in \Xi$.

Ontological models where outcome determinism does not hold are called outcome indeterministic. Of the class of outcome-indeterministic ontological models, the ones that are related to outcome-deterministic models via Fine's theorem are the models that satisfy *factorizability*.

Factorizability is the assumption that for every jointly measurable set of measurements $\{M_s | s \in S\}$, the response function for every outcome of a joint measurement M_S is the product of the response functions of measurements in the jointly measurable set: $\xi(k_S|M_S, \lambda) = \prod_{s \in S} \xi(k_s|M_s, \lambda)$. Note that $k_S \in \mathcal{K}_{M_S}$ and $k_s \in \mathcal{K}_{M_s}$, where $\mathcal{K}_{M_S}$ is the Cartesian product of the outcome sets $\mathcal{K}_{M_s}, s \in S$.

In the next section I will point out how these assumptions are related via Fine's theorem. This will be followed by a discussion of how, although factorizability is a physically motivated assumption in locally causal models, it does not admit such a motivation in the more general case of noncontextual models. Fine's theorem thus serves to delineate a mathematical boundary between KS-noncontextual models and noncontextual models which are not factorizable.

III. FINE'S THEOREM FOR NONCONTEXTUAL MODELS

Theorem 1. Given a set of measurements $\{M_1, \ldots, M_N\}$ with jointly measurable subsets $S \subset \{1, \ldots, N\}$, where each measurement $M_s, s \in S$, takes values labeled by $k_s \in \mathcal{K}_{M_s}$, the following propositions are equivalent:

(1) For a given preparation $P \in \mathcal{P}$ of the system there exists a joint probability distribution $p(k_1, \ldots, k_N|P)$ that recovers the marginal statistics for jointly measurable subsets predicted by the operational theory (such as quantum theory) under consideration, i.e., $\forall S \subset \{1, \ldots, N\}$, $p(k_S|M_S; P) = \sum_{k_i : i \notin S} p(k_1, \ldots, k_N|P)$, where $k_S \in \mathcal{K}_{M_S}$.

(2) There exists a measurement-noncontextual and outcome-deterministic, i.e., KS-noncontextual, model for these measurements.

(3) There exists a measurement-noncontextual and factorizable model for these measurements.

Proof. The proof of equivalence of the three propositions proceeds as follows: Proposition 3 ⇒ Proposition 1, Proposition 1 ⇒ Proposition 2, Proposition 2 ⇒ Proposition 3.

Proposition 3 ⇒ Proposition 1

By Proposition 3, the assumption of measurement noncontextuality requires that the single-measurement response functions in the model be of the form $\xi(k_i|M_i; \lambda) \in [0,1]$, so that each response function is independent of the contexts—jointly measurable subsets—that the corresponding measurement may be a part of. Of course, the assumption of measurement noncontextuality only applies once it is verified that for any $P \in \mathcal{P}$ the operational statistics $p(k_i|M_i; P)$ of measurement M_i is the same across all the jointly measurable subsets in which it appears. The response function is therefore conditioned only by M_i and the ontic state λ associated with the system. Proposition 3 requires, in addition, factorizability, i.e., for all jointly measurable subsets $S \subset \{1, \ldots, N\}$, $\xi(k_S|M_S; \lambda) = \prod_{s \in S} \xi(k_s|M_s; \lambda)$. Factorizability amounts to the assumption that the correlations between measurement outcomes are established only via the ontic state of the system—the measurements do not "talk" to each other except via λ. Now define

$$\xi(k_1, \ldots, k_N|\lambda) \equiv \prod_{i=1}^{N} \xi(k_i|M_i; \lambda), \tag{6}$$

so that marginalizing this distribution over k_i, $i \notin S$, yields $\xi(k_S|M_S; \lambda)$ for every jointly measurable subset $S \subset \{1, \ldots, N\}$.

Assuming the ontological model reproduces the operational statistics, there must exist a probability density function $\mu(\lambda|P)$ for any $P \in \mathcal{P}$, such that

$$\int d\lambda \, \xi(k_S|M_S; \lambda) \mu(\lambda|P) = p(k_S|M_S; P). \tag{7}$$

Then define

$$p(k_1 \cdots k_N|P) \equiv \int d\lambda \, \xi(k_1 \cdots k_N|\lambda) \mu(\lambda|P), \tag{8}$$

which marginalizes on k_S to

$$p(k_S|P) = \sum_{k_i : i \notin S} p(k_1 \cdots k_N|P) \tag{9}$$

$$= \int d\lambda \sum_{k_i : i \notin S} \xi(k_1 \cdots k_N|\lambda) \mu(\lambda|P) \tag{10}$$

$$= \int d\lambda \, \xi(k_S|M_S; \lambda) \mu(\lambda|P) \tag{11}$$

$$= p(k_S|M_S; P). \tag{12}$$

Thus, Proposition 3 $\Rightarrow$ Proposition 1.

Proposition 1 $\Rightarrow$ Proposition 2

By Proposition 1, for a given $P \in \mathcal{P}$ there exists a $p(k_1 \cdots k_N | P)$ such that $p(k_S | M_S; P) = \sum_{k_i : i \notin S} p(k_1 \cdots k_N | P)$, for all jointly measurable subsets $S \subset \{1, \ldots, N\}$. Now there exists a probability density function $\mu(\lambda | P)$ such that

$$p(k_1 \cdots k_N | P) = \int d\lambda \xi(k_1 \cdots k_N | \lambda) \mu(\lambda | P), \quad (13)$$

where $\xi(k_1 \cdots k_N | \lambda) \in \{0, 1\}$. This is possible because any probability distribution can be decomposed as a convex sum over deterministic distributions. Also, $p(k_j | M_j; P) = \sum_{k_i : i \neq j} p(k_1, \ldots, k_N | P)$, so

$$p(k_j | M_j; P) = \int d\lambda \mu(\lambda | P) \sum_{k_i : i \neq j} \xi(k_1 \cdots k_N | \lambda), \quad (14)$$

which allows the definition

$$\xi(k_j | M_j; \lambda) \equiv \sum_{k_i : i \neq j} \xi(k_1 \cdots k_N | \lambda) \in \{0, 1\}, \forall j \in \{1 \cdots N\}. \quad (15)$$

Since these are deterministic distributions,

$$\xi(k_1 \cdots k_N | \lambda) = \prod_{j=1}^{N} \xi(k_j | M_j; \lambda). \quad (16)$$

Finally,

$$p(k_S | M_S; P) = \int d\lambda \mu(\lambda | P) \prod_{s \in S} \xi(k_s | M_s; \lambda), \quad (17)$$

so there exists a measurement-noncontextual and outcome-deterministic model, i.e., Proposition 1 $\Rightarrow$ Proposition 2.

Proposition 2 $\Rightarrow$ Proposition 3

By Proposition 2, $\xi(k_i | M_i; \lambda) \in \{0, 1\}, \forall i \in \{1 \cdots N\}$, such that

$$p(k_S | M_S; P) = \int d\lambda \mu(\lambda | P) \prod_{s \in S} \xi(k_s | M_s; \lambda), \quad (18)$$

$\forall$ jointly measurable subsets $S \subset \{1 \cdots N\}$. Clearly this model is also a measurement-noncontextual and factorizable model because the assumption of outcome determinism implies factorizability:

$$\xi(k_S | M_S; \lambda) = \prod_{s \in S} \xi(k_s | M_s; \lambda). \quad (19)$$

∎

Note that this theorem itself is not new, but this particular reading of it in the framework of generalized noncontextuality *is* new. In particular, the purpose of this restatement is to highlight why outcome determinism is not an assumption that can be taken for granted in noncontextual ontological models. Versions of this theorem have appeared in the literature following Fine's original insight [6,7]. The fact that Proposition 2 implies Proposition 1 has been shown earlier in Ref. [10]. A similar result in the language of sheaf theory can be found Ref. [14], where the authors point out factorizability as the underlying assumption in Bell-local and KS-noncontextual models: in effect they show the equivalence of Proposition 1

and Proposition 3. The sense in which Ref. [14] refers to "noncontextuality" is the notion of KS noncontextuality, and while it is possible to provide a unified account of Bell locality and KS noncontextuality at a mathematical level, the generalized notion of noncontextuality [2] does not admit such an account. In particular, their definition of noncontextuality is stronger than the Spekkens' definition of measurement non-contextuality. Indeed, generalized noncontextuality subsumes KS noncontextuality but is not equivalent to it.

Fine's theorem for Bell scenarios: Translating the preceding notions from noncontextual models to Bell-local models amounts to replacing measurement noncontextual and out-come deterministic by locally deterministic and measurement noncontextual and factorizable by locally causal. Consider the case of two-party Bell scenarios for simplicity, although the same considerations extend to general multiparty Bell scenarios in a straightforward manner. A two-party Bell scenario consists of measurements $\{M_1, \ldots, M_N\}$, where $\{M_1, \ldots, M_n\}, n < N - 1$, are the measurement settings available to one party, say Alice, and $\{M_{n+1}, \ldots, M_N\}$ are the measurement settings available to the other party, say Bob. The outcomes are denoted by $k_i \in \mathcal{K}_{M_i}$ for the respective measurement settings M_i. The jointly measurable subsets are given by $S \in \{\{i, j\} | i \in \{1, \ldots, n\}, j \in \{n+1, \ldots, N\}\}$. Bell's assumption of local causality captures the notion of a measurement noncontextual and factorizable model:

$$p(k_S | M_S; P) = p(k_i, k_j | M_i, M_j; P) \quad (20)$$

$$= \int d\lambda \mu(\lambda | P) \xi(k_i | M_i, \lambda) \xi(k_j | M_j, \lambda). \quad (21)$$

Once factorizability is justified from Bell's assumption of local causality in this manner, Fine's theorem ensures that—so far as the existence of hidden variable models is concerned—it is irrelevant whether the response functions for the measurement outcomes are deterministic or indeterministic. One does not need to worry about whether outcome determinism for measurements is justified in Bell scenarios precisely because factorizability along with Fine's theorem absolves one of the need to provide such a justification. The crucial point, then, is the validity of factorizability in the more general case of noncontextual models. In general, factorizability is not justified in noncontextual models and, following Spekkens, one must distinguish between the issue of noncontextuality and that of outcome determinism when considering ontological models of an operational theory [2]. If the goal is—as it should be—to obtain an experimental test of noncontextual models independent of the truth of quantum theory, then one needs to derive noncontextuality inequalities that do not rely on outcome determinism at all. This is because Fine's theorem for noncontextual models is of limited applicability—namely, outcome-indeterministic response functions which satisfy factorizability are shown by it to achieve no more generality than is already captured by outome-deterministic response functions in a KS-noncontextual model. Outcome-indeterministic response functions that do not satisfy factorizability are not taken into account in a KS-noncontextual model.

For ontological models of operational *quantum* theory, outcome determinism for sharp (projective) measurements can be shown to follow from the assumption of preparation

noncontextuality [2]. Such a justification is not available for unsharp (nonprojective) measurements, which should therefore be represented by outcome-indeterministic response functions. This issue has been discussed at length by Spekkens and the reader is referred to Ref. [13] for why and how this must be so. Therefore, to consider noncontextuality for unsharp measurements in full generality the noncontextuality inequalities of interest are those which do not assume factorizability. An example is the LSW inequality for Specker's scenario [10] that does not rely on factorizability, although it does use the assumption of outcome determinism for sharp (projective) measurements. The LSW inequality has been shown to be violated by quantum predictions [11], thus ruling out noncontextual models of quantum theory without invoking factorizability. Note that the distinction between sharp and unsharp measurements is not part of the definition of a Bell-local model and one never has to worry about this distinction to derive Bell's theorem. This distinction, however, becomes relevant for noncontextual models of quantum theory, where the words "sharp" and "unsharp" have a clear meaning, the former referring to projective measurements and the latter to nonprojective measurements.

In the next section, the polytope of correlations admissible in Specker's scenario is characterized.

IV. CORRELATIONS IN SPECKER'S SCENARIO

In this section three noncontextuality inequalities relevant to the correlations in Specker's scenario are derived. They are shown to be equivalent to the known LSW inequality under relabeling of measurement outcomes. This scenario involves three binary measurements $\{M_1, M_2, M_3\}$, which are pairwise jointly measurable with outcomes labeled by $X_i \in \{0,1\}$ for $i \in \{1,2,3\}$. The statistics involved in Specker's scenario for a given preparation $P \in \mathcal{P}$ can be understood as a set of 12 probabilities, four for each pairwise joint measurement M_{ij},

$$\mathcal{S} \equiv \{p(X_i X_j | M_{ij}; P) | X_i, X_j \in \{0,1\}, i, j \in \{1,2,3\}, i < j\},$$
(22)

subject to the obvious constraints of positivity,

$$p(X_i X_j | M_{ij}; P) \geqslant 0 \quad \forall X_i, X_j, M_{ij},$$
(23)

and normalization,

$$\sum_{X_i, X_j} p(X_i X_j | M_{ij}; P) = 1 \quad \forall M_{ij}.$$
(24)

In addition to positivity and normalization, the statistics is assumed to obey the following condition:

$$\sum_{X_j} p(X_i X_j | M_{ij}; P) = \sum_{X_k} p(X_i X_k | M_{ik}; P)$$
(25)

$$\equiv p(X_i | M_i; P),$$
(26)

for all $i < j,k$ where $i,j,k \in \{1,2,3\}$. Denoting $\sum_{X_j} p(X_i X_j | M_{ij}; P) \equiv p(X_i | M_i^j; P)$, and $\sum_{X_k} p(X_i X_k | M_{ik}; P) \equiv p(X_i | M_i^k; P)$, the condition becomes

$$p(X_i | M_i^j; P) = p(X_i | M_i^k; P) \equiv p(X_i | M_i; P).$$
(27)

That is, the statistics of M_i^j, which is obtained by marginalizing the statistics of joint measurement M_{ij}, is identical to the statistics of M_i^k, which is obtained by marginalizing the statistics of joint measurement M_{ik}. If what has been measured is indeed a unique observable M_i then its statistics relative to any preparation $P \in \mathcal{P}$ should remain the same across joint measurements with different observables M_j and M_k. Failure to meet this condition implies a failure of joint measurability: then one can distinguish between M_i^j and M_i^k from their statistics relative to some preparation and they would therefore correspond to two different marginal observables M_i^j and M_i^k rather than a unique observable M_i. This condition is often called the *no-disturbance* condition in the literature on contextuality. Operational quantum theory obeys the no-disturbance condition for joint measurements of generalized observables (which need not be projective or sequential).

A. Kochen-Specker (KS) inequalities for Specker's scenario

The four necessary and sufficient inequalities characterizing correlations which admit a KS-noncontextual model in Specker's scenario are given by

$$R_3 \equiv p(X_1 \neq X_2 | M_{12}, P) + p(X_2 \neq X_3 | M_{23}, P)$$
$$+ p(X_1 \neq X_3 | M_{13}, P) \leqslant 2$$
(28)

and

$$R_0 \equiv p(X_1 \neq X_2 | M_{12}, P) - p(X_2 \neq X_3 | M_{23}, P)$$
$$- p(X_1 \neq X_3 | M_{13}, P) \leqslant 0,$$
(29)

$$R_1 \equiv p(X_2 \neq X_3 | M_{23}, P) - p(X_1 \neq X_3 | M_{13}, P)$$
$$- p(X_1 \neq X_2 | M_{12}, P) \leqslant 0,$$
(30)

$$R_2 \equiv p(X_1 \neq X_3 | M_{13}, P) - p(X_1 \neq X_2 | M_{12}, P)$$
$$- p(X_2 \neq X_3 | M_{23}, P) \leqslant 0.$$
(31)

These inequalities have earlier appeared in Ref. [15]. A derivation is provided in Appendix B. Furthermore, these inequalities exhibit a curious property that no two of them can be violated by the same set of experimental statistics.

Lemma 1. There exists no set of distributions $\{p(X_i, X_j | M_{ij}, P) | (ij) \in \{(12),(23),(13)\}\}$ that can violate any two of the four KS inequalities simultaneously.

Proof. Denoting $w_{12} \equiv p(X_1 \neq X_2 | M_{12}, P)$, $w_{23} \equiv p(X_2 \neq X_3 | M_{23}, P)$, and $w_{13} \equiv p(X_1 \neq X_3 | M_{13}, P)$, the four KS inequalities can be rewritten as

$$R_3 \equiv w_{12} + w_{23} + w_{13} \leqslant 2,$$
(32)

$$R_0 \equiv w_{12} - w_{23} - w_{13} \leqslant 0,$$
(33)

$$R_1 \equiv w_{23} - w_{13} - w_{12} \leqslant 0,$$
(34)

$$R_2 \equiv w_{13} - w_{23} - w_{12} \leqslant 0.$$
(35)

Now, violation of each of these is equivalent to the following, since $0 \leqslant w_{12}, w_{23}, w_{13} \leqslant 1$:

$$R_3 > 2 \Leftrightarrow w_{12} + w_{23} + w_{13} > 2,$$

$$R_0 > 0 \Leftrightarrow w_{12} > w_{23} + w_{13} \Rightarrow w_{12} + w_{23} + w_{13} < 2,$$

$$R_1 > 0 \Leftrightarrow w_{23} > w_{12} + w_{13} \Rightarrow w_{12} + w_{23} + w_{13} < 2$$
and $w_{12} < w_{23} - w_{13},$

$$R_2 > 0 \Leftrightarrow w_{13} > w_{12} + w_{23} \Rightarrow w_{12} + w_{23} + w_{13} < 2$$
and $w_{12} < w_{13} - w_{23}.$

It follows that violation of each inequality above is in conflict with a violation of each of the other three inequalities. Hence, there exist no conceivable measurement statistics that violate any two of the four KS inequalities simultaneously. ∎

B. Noncontextuality (NC) inequalities for Specker's scenario

Consider the *predictability* of each measurement M_k defined as

$$\eta_{M_k} \equiv \max_P \left\{ 2 \max_{X_k} p(X_k | M_k, P) - 1 \right\}, \qquad (36)$$

where P is any preparation of the system. Assuming the three measurements in Specker's scenario have the same predictability $\eta_0 \equiv \eta_{M_1} = \eta_{M_2} = \eta_{M_3}$, the following noncontextuality inequalities hold.

1. LSW inequality

$$R_3 = p(X_1 \neq X_2 | M_{12}, P) + p(X_2 \neq X_3 | M_{23}, P)$$
$$+ p(X_1 \neq X_3 | M_{13}, P) \leqslant 3 - \eta_0, \qquad (37)$$

2. Three more inequalities

$$R_0 = p(X_1 \neq X_2 | M_{12}, P) - p(X_2 \neq X_3 | M_{23}, P)$$
$$- p(X_1 \neq X_3 | M_{13}, P) \leqslant 1 - \eta_0, \qquad (38)$$

$$R_1 = p(X_2 \neq X_3 | M_{23}, P) - p(X_1 \neq X_3 | M_{13}, P)$$
$$- p(X_1 \neq X_2 | M_{12}, P) \leqslant 1 - \eta_0, \qquad (39)$$

$$R_2 = p(X_1 \neq X_3 | M_{13}, P) - p(X_1 \neq X_2 | M_{12}, P)$$
$$- p(X_2 \neq X_3 | M_{23}, P) \leqslant 1 - \eta_0. \qquad (40)$$

These inequalities are derived in Appendix B. Note that violation of each of these inequalities implies the violation of the corresponding KS inequalities (recovered for $\eta_0 = 1$), but not conversely.

Lemma 2. There exists no set of distributions $\{p(X_i, X_j | M_{ij}, P) | (ij) \in \{(12), (23), (13)\}\}$ that can violate any two of the four NC inequalities simultaneously.

Proof. The proof trivially follows from Lemma 1, since violation of any NC inequality implies violation of the corresponding KS inequality. ∎

The predictability η_0 quantifies how predictable a measurement can be made in a variation over preparations: KS

inequalities make sense only when $\eta_0 = 1$, i.e., it is possible to find a preparation which makes a given measurement perfectly predictable, a condition which is naturally satisfied by sharp (projective) measurements in quantum theory. For the case of unsharp measurements, $\eta_0 < 1$, and the noncontextuality inequalities take this into account. When $\eta_0 = 0$, that is, when the measurement outcomes are uniformly random (or completely unpredictable), the upper bounds in the noncontextuality inequalities become trivial and a noncontextual model is always possible: simply ignore the system and toss a fair coin to decide whether to output $(X_i = 0, X_j = 1)$ or $(X_i = 1, X_j = 0)$ when a pair of measurements $\{M_i, M_j\}$ is jointly implemented,

$$p(X_i, X_j | M_{ij}, P) = \tfrac{1}{2} (\delta_{X_i,0} \delta_{X_j,1} + \delta_{X_i,1} \delta_{X_j,0}). \qquad (41)$$

Clearly, $R_3 = 3$ for this, and $\eta_0 = 0$ since the marginal for each measurement M_i is uniformly random independent of the preparation, so the LSW inequality cannot be violated. This admits a noncontextual model since the response function for each measurement M_i is a fair coin flip independent of the system's ontic state and also of which other measurement it is jointly implemented with. The key feature that the LSW inequality captures is this: that it is not possible to have a high degree of anticorrelation R_3 and a high degree of predictability η_0 in a noncontextual model, and that there is a tradeoff between the two, given here by $R_3 + \eta_0 \leqslant 3$. Contextuality in this sense signifies the ability to generate (anti)correlations which violate this tradeoff for values of $\eta_0 < 1$: the case $\eta_0 = 1$, as mentioned, is already covered by the usual KS inequalities, and for $\eta_0 = 0$ there is no nontrivial tradeoff imposed by noncontextual models.

C. Equivalence under relabeling of measurement outcomes

The four NC inequalities (also the KS inequalities) are equivalent under relabeling measurement outcomes: To go from $R_3 \leqslant 3 - \eta_0$ to $R_0 \leqslant 1 - \eta_0$, simply relabel the measurement outcomes of M_3 as $X_3 \to X_3' = 1 - X_3$, so that after the relabeling (denoted by primed quantities): $w_{12}' = w_{12}, w_{23}' = 1 - w_{23}, w_{13}' = 1 - w_{13}$, and $R_3' \equiv w_{12}' + w_{23}' + w_{13}' \leqslant 3 - \eta_0$ becomes $w_{12} + (1 - w_{23}) + (1 - w_{13}) \leqslant 3 - \eta_0$ which can be rewritten as $R_3' = R_0 = w_{12} - w_{23} - w_{13} \leqslant 1 - \eta_0$. Similarly, relabeling measurement outcomes of M_2 takes $R_3 \leqslant 3 - \eta_0$ to $R_2 \leqslant 1 - \eta_0$ and relabeling measurement outcomes of M_1 takes $R_3 \leqslant 3 - \eta_0$ to $R_1 \leqslant 1 - \eta_0$.

D. Quantum violation of noncontextuality inequalities for Specker's scenario

Quantum realization of Specker's scenario involves three unsharp qubit POVMs $M_k = \{E_0^k, E_1^k\}, k \in \{1, 2, 3\}$, where the effects are given by

$$E_{X_k}^k \equiv \frac{1}{2} I + (-1)^{X_k} \frac{\eta}{2} \vec{\sigma} \cdot \hat{n}_k, \quad X_k \in \{0, 1\}, 0 \leqslant \eta \leqslant 1. \qquad (42)$$

These can be rewritten as

$$E_{X_k}^k = \eta \Pi_{X_k}^k + (1 - \eta) \frac{I}{2}, \qquad (43)$$

where $\Pi_{X_k}^k = \frac{1}{2}[I + (-1)^{X_k}\vec{\sigma} \cdot \hat{n}_k]$ are the corresponding projectors. That is, M_k is a noisy version of the projective measurement of spin along the $\hat{n}_k$ direction, where the sharpness of the POVM is given by η. In this case, $p(X_k|M_k, P) = \text{Tr}(\rho_P E_{X_k}^k)$, where ρ_P is the density matrix for preparation P of the system and the predictability can be easily shown to be η: the preparation maximizing η_{M_k} is a pure state along the $\hat{n}_k$ axis, i.e., $\rho_P = \Pi_{X_k}^k$.

Quantum violation of the LSW inequality has already been shown in Ref. [11]. On account of the equivalence of the four NC inequalities under relabeling of measurement outcomes, the violation of the other three NC inequalities besides LSW follows from appropriate relabelings of measurement outcomes in the quantum violation demonstrated in Ref. [11].

E. Specker polytope

The statistics allowed in Specker's scenario, given that the no-disturbance condition holds, can be understood as a convex polytope in $\mathbb{R}^6$ with 12 extreme points or vertices, eight of which are deterministic and four indeterministic. The measurement statistics are given by the vector of 12 probabilities $\vec{v}(P) = [v_{X_i X_j}^{ij}(P)|X_i, X_j \in \{0,1\}, i,j \in \{1,2,3\}, i < j]$, where $v_{X_i X_j}^{ij}(P) \equiv p(X_i X_j|M_{ij}; P)$, constrained by the positivity, normalization, and no-disturbance conditions which reduce the number of independent probabilities in $\vec{v}(P)$ from 12 to 6.

The deterministic vertices, which admit KS-noncontextual models, correspond to the eight possible tripartite joint distributions of the form $p(X_1, X_2, X_3|P) \equiv \delta_{X_1, X_1(P)}\delta_{X_2, X_2(P)}, \delta_{X_3, X_3(P)}$, where $X_1(P), X_2(P), X_3(P) \in \{0,1\}$. The deterministic vertex $\vec{v}(P)$ can be obtained from this joint distribution as $v_{X_i X_j}^{ij}(P) = \sum_{X_k, k \neq i,j} p(X_1, X_2, X_3|P) = \delta_{X_i, X_i(P)}\delta_{X_j, X_j(P)}$. These vertices are labeled lexicographically, $(X_1(P), X_2(P), X_3(P))$ as the decimal equivalent of binary number $X_1(P)X_2(P)X_3(P)$:

$$\vec{v}_0(P) : v_{00}^{12}(P) = v_{00}^{23}(P) = v_{00}^{13}(P) = 1, \tag{44}$$

$$\vec{v}_1(P) : v_{00}^{12}(P) = v_{01}^{23}(P) = v_{01}^{13}(P) = 1, \tag{45}$$

$$\vec{v}_2(P) : v_{01}^{12}(P) = v_{10}^{23}(P) = v_{00}^{13}(P) = 1, \tag{46}$$

$$\vec{v}_3(P) : v_{01}^{12}(P) = v_{11}^{23}(P) = v_{01}^{13}(P) = 1, \tag{47}$$

$$\vec{v}_4(P) : v_{10}^{12}(P) = v_{00}^{23}(P) = v_{10}^{13}(P) = 1, \tag{48}$$

$$\vec{v}_5(P) : v_{10}^{12}(P) = v_{01}^{23}(P) = v_{11}^{13}(P) = 1, \tag{49}$$

$$\vec{v}_6(P) : v_{11}^{12}(P) = v_{10}^{23}(P) = v_{10}^{13}(P) = 1, \tag{50}$$

$$\vec{v}_7(P) : v_{11}^{12}(P) = v_{11}^{23}(P) = v_{11}^{13}(P) = 1. \tag{51}$$

Note that these deterministic vertices satisfy all the four KS inequalities, Eqs. (28)–(31), and therefore also the four noncontextuality inequalities, Eqs. (37)–(40). That is, they admit a KS-noncontextual model. Indeed, the convex set that these eight extreme points define is a KS-noncontextuality polytope, analogous to a Bell polytope in a Bell scenario. This polytope is a subset of the larger Specker polytope

which in addition to these eight vertices includes the four indeterministic vertices in Specker's scenario.

The indeterministic vertices, which do *not* admit KS-noncontextual models, correspond to the four sets of pairwise joint distributions given by

$$\vec{v}_8(P) : \tag{52}$$

$$v_{01}^{12}(P) = v_{10}^{12}(P) = \tfrac{1}{2}, \tag{53}$$

$$v_{00}^{23}(P) = v_{11}^{23}(P) = \tfrac{1}{2}, \tag{54}$$

$$v_{00}^{13}(P) = v_{11}^{13}(P) = \tfrac{1}{2}, \tag{55}$$

$$\vec{v}_9(P) : \tag{56}$$

$$v_{00}^{12}(P) = v_{11}^{12}(P) = \tfrac{1}{2}, \tag{57}$$

$$v_{01}^{23}(P) = v_{10}^{23}(P) = \tfrac{1}{2}, \tag{58}$$

$$v_{00}^{13}(P) = v_{11}^{13}(P) = \tfrac{1}{2}, \tag{59}$$

$$\vec{v}_{10}(P) : \tag{60}$$

$$v_{00}^{12}(P) = v_{11}^{12}(P) = \tfrac{1}{2}, \tag{61}$$

$$v_{00}^{23}(P) = v_{11}^{23}(P) = \tfrac{1}{2}, \tag{62}$$

$$v_{01}^{13}(P) = v_{10}^{13}(P) = \tfrac{1}{2}, \tag{63}$$

$$\vec{v}_{11}(P) : \tag{64}$$

$$v_{01}^{12}(P) = v_{10}^{12}(P) = \tfrac{1}{2}, \tag{65}$$

$$v_{01}^{23}(P) = v_{10}^{23}(P) = \tfrac{1}{2}, \tag{66}$$

$$v_{01}^{13}(P) = v_{10}^{13}(P) = \tfrac{1}{2}. \tag{67}$$

The vertex $\vec{v}_8(P)$ violates inequalities (29) and (38) ($\eta_0 > 0$), $\vec{v}_9(P)$ violates inequalities (30) and (39) ($\eta_0 > 0$), $\vec{v}_{10}(P)$ violates inequalities (31) and (40) ($\eta_0 > 0$), and $\vec{v}_{11}(P)$ violates inequalities (28) and (37) ($\eta_0 > 0$). Note that these vertices are equivalent under relabelings, that is, $\vec{v}_8(P)$ turns to $\vec{v}_{11}(P)$ on relabeling outcomes of M_3, $\vec{v}_9(P)$ to $\vec{v}_{11}(P)$ on relabeling outcomes of M_1, and $\vec{v}_{10}(P)$ to $\vec{v}_{11}(P)$ on relabeling outcomes of M_2. Note that the vertex $\vec{v}_{11}(P)$ corresponds to the "overprotective seer" (OS) correlations of Ref. [10] which maximally violate the LSW inequality when $\eta_0 < 1$.

F. Limitations of the joint probability distribution criterion for deciding contextuality

All the Bell-Kochen-Specker type analyses of contextuality ultimately hinge on ruling out the existence of a joint probability distribution that reproduces the operational statistics of various jointly measurable observables as marginals. Deciding whether such a joint distribution exists is called a marginal problem [16]. That this is a limited criterion to decide the

question of contextuality without also making the assumption of outcome determinism or factorizability is borne out by correlations in Specker's scenario that lie outside the polytope of correlations admissible in KS-noncontextual models but are realizable in noncontextual models. Violation of the LSW inequality by unsharp measurements in quantum theory rules out such noncontextual models [11].

Once outcome determinism for unsharp measurements (ODUM, cf. [13]) is abandoned, the existence of a joint distribution is no longer necessary to characterize noncontextual models. Furthermore, in the case of an arbitrary operational theory which is not quantum theory it is not obvious whether outcome determinism for measurements can at all be justified from the assumption of preparation and measurement noncontextuality. An experimentally interesting and robust noncontextuality inequality should not assume that the operational theory describing the experiment is quantum theory and instead derive from the assumption of noncontextuality alone, given some operational equivalences between preparation procedures or measurement procedures. Violation of the LSW inequality only indicates that *quantum theory* does not admit a noncontextual ontological model. The ideal to aspire for is something akin to Bell inequalities which are theory independent. That such an ideal is achievable will be shown in a forthcoming paper [17].

V. CONCLUSION

To summarize, the chief takeaways from this paper are the following.

(1) Fine's theorem for noncontextual models only applies in cases where the correlations between measurement outcomes are mediated exclusively by the ontic state λ of the system. When this is not the case and factorizability fails, it is possible that the measurement outcomes share correlations that are not on account of the measured system but an artifact of the measurement apparatus. Considering noncontextual models which are not factorizable allows one to handle this situation.

(2) The no-disturbance polytope of Specker's scenario admits four indeterministic extremal points, related to each other by relabelings of measurement outcomes, that are related to the "OS box" of Ref. [10]. Corresponding to these four extremal points are four Kochen-Specker inequalities assuming outcome determinism, and four noncontextuality inequalities that do not assume outcome determinism.

All this points out the need to further investigate how a failure of outcome determinism or factorizability in the case of more well-known KS inequalities should be handled. Another open question is how to derive noncontextuality inequalities for arbitrary operational theories, rather than just quantum theory, without any assumption of outcome determinism or factorizability. These questions will be taken up in future work.

APPENDIX A: CONSTRAINTS ON THE OPERATIONAL STATISTICS FROM NORMALIZATION AND NO-DISTURBANCE

The notation here is simplified as follows: the measurements are denoted by $e \equiv M_1, f \equiv M_2, g \equiv M_3$, and their outcomes by $e_k \equiv (X_1 = k), f_k \equiv (X_2 = k), g_k \equiv (X_3 = k)$, where $k \in \{0,1\}$. Thus there are three binary observables, e, f, g, each taking values in $\{0,1\}$ and measured on a system prepared according to some preparation P. e_0 denotes the outcome $e = 0$ and e_1 denotes $e = 1$. Analogous notation applies for outcomes of f and g as well. The probability distributions on these observables associated with the preparation P are denoted by $w_P(e) \equiv \{w_P(e_0), w_P(e_1)\}, w_P(f) \equiv \{w_P(f_0), w_P(f_1)\}, w_P(g) \equiv \{w_P(g_0), w_P(g_1)\}$. The experimental statistics correspond to the joint measurement of every pair of observables:

$$w_P(e,f) \equiv \{w_P(e_0,f_0), w_P(e_0,f_1), w_P(e_1,f_0), w_P(e_1,f_1)\},$$
$$w_P(f,g) \equiv \{w_P(f_0,g_0), w_P(f_0,g_1), w_P(f_1,g_0), w_P(f_1,g_1)\},$$
$$w_P(e,g) \equiv \{w_P(e_0,g_0), w_P(e_0,g_1), w_P(e_1,g_0), w_P(e_1,g_1)\}.$$

In addition to the usual positivity and normalization constraints for probability distributions, the no-disturbance condition on the pairwise joint distributions yields

$$w_P(e_0,f_0) + w_P(e_0,f_1) = w_P(e_0,g_0) + w_P(e_0,g_1)$$
$$\equiv w_P(e_0)$$
$$\Rightarrow w_P(e_1,f_0) + w_P(e_1,f_1) = w_P(e_1,g_0) + w_P(e_1,g_1)$$
$$\equiv w_P(e_1),$$
$$w_P(f_0,g_0) + w_P(f_0,g_1) = w_P(e_0,f_0) + w_P(e_1,f_0)$$
$$\equiv w_P(f_0)$$
$$\Rightarrow w_P(f_1,g_0) + w_P(f_1,g_1) = w_P(e_0,f_1) + w_P(e_1,f_1)$$
$$\equiv w_P(f_1),$$
$$w_P(e_0,g_0) + w_P(e_1,g_0) = w_P(f_0,g_0) + w_P(f_1,g_0)$$
$$\equiv w_P(g_0)$$
$$\Rightarrow w_P(e_0,g_1) + w_P(e_1,g_1) = w_P(f_0,g_1) + w_P(f_1,g_1)$$
$$\equiv w_P(g_1).$$

Normalization gets rid of three parameters out of the 12 in the experimental statistics while no-disturbance eliminates three more parameters. There are, therefore, six independent parameters describing the experimental statistics:

$$w_{12} = w_P(e_0,f_1) + w_P(e_1,f_0), \tag{A1}$$

$$w_{23} = w_P(f_0,g_1) + w_P(f_1,g_0), \tag{A2}$$

$$w_{13} = w_P(e_0,g_1) + w_P(e_1,g_0), \tag{A3}$$

$$p_1 \equiv w_P(e_0), \tag{A4}$$

$$p_2 \equiv w_P(f_0), \tag{A5}$$

$$p_3 \equiv w_P(g_0), \tag{A6}$$

subject to $0 \leqslant w_{12}, w_{23}, w_{13}, p_1, p_2, p_3 \leqslant 1$. Using the no-disturbance and normalization conditions:

$$w_P(e_0, f_1) = \frac{w_{12} + p_1 - p_2}{2},$$

$$w_P(e_1, f_0) = \frac{w_{12} - p_1 + p_2}{2},$$

$$w_P(e_0, f_0) = \frac{p_1 + p_2 - w_{12}}{2},$$

$$w_P(e_1, f_1) = 1 - \frac{w_{12} + p_1 + p_2}{2},$$

$$w_P(f_0, g_1) = \frac{w_{23} + p_2 - p_3}{2},$$

$$w_P(f_1, g_0) = \frac{w_{23} - p_2 + p_3}{2},$$

$$w_P(f_0, g_0) = \frac{p_2 + p_3 - w_{23}}{2},$$

$$w_P(f_1, g_1) = 1 - \frac{w_{23} + p_2 + p_3}{2},$$

$$w_P(e_0, g_1) = \frac{w_{13} + p_1 - p_3}{2},$$

$$w_P(e_1, g_0) = \frac{w_{13} - p_1 + p_3}{2},$$

$$w_P(e_0, g_0) = \frac{p_1 + p_3 - w_{13}}{2},$$

$$w_P(e_1, g_1) = 1 - \frac{w_{13} + p_1 + p_3}{2}.$$

The positivity requirements on these translate to the following inequalities:

$$|p_1 - p_2| \leqslant w_{12} \leqslant p_1 + p_2 \leqslant 2 - w_{12}, \tag{A7}$$

$$|p_2 - p_3| \leqslant w_{23} \leqslant p_2 + p_3 \leqslant 2 - w_{23}, \tag{A8}$$

$$|p_1 - p_3| \leqslant w_{13} \leqslant p_1 + p_3 \leqslant 2 - w_{13}. \tag{A9}$$

APPENDIX B: DERIVING THE KS AND NC INEQUALITIES

1. KS inequalities

The KS inequalities derive from the existence of a joint probability distribution $p(X_1 X_2 X_3)$ such that $p(X_i X_j | M_{ij}, P) = \sum_{X_k} p(X_1 X_2 X_3)$, where i, j, k are distinct indices in $\{1, 2, 3\}$. Therefore the following must hold:

$$p(001) = p(00|M_{12}, P) - p(000),$$

$$p(010) = p(00|M_{13}, P) - p(000),$$

$$p(100) = p(00|M_{23}, P) - p(000),$$

$$p(011) = p(01|M_{12}, P) - p(010)$$
$$= p(01|M_{12}, P) - p(00|M_{13}, P) + p(000),$$

$$p(101) = p(10|M_{12}, P) - p(100)$$
$$= p(10|M_{12}, P) - p(00|M_{23}, P) + p(000),$$

$$p(110) = p(10|M_{13}, P) - p(100)$$
$$= p(10|M_{13}, P) - p(00|M_{23}, P) + p(000),$$

$$p(111) = 1 - p(00|M_{12}, P) - p(01|M_{12}, P) - p(10|M_{12}, P)$$
$$- p(10|M_{13}, P) + p(00|M_{23}, P) - p(000).$$

Expressing the probabilities in terms of the six free parameters identified earlier, namely, the anticorrelation probabilities w_{12}, w_{23}, w_{13}, and the marginals p_1, p_2, p_3, the positivity constraints $0 \leqslant p(X_1 X_2 X_3) \leqslant 1$ require

$$0 \leqslant p(000) \leqslant 1,$$

$$0 \leqslant p(001) \leqslant 1$$
$$\Leftrightarrow p_1 + p_2 - 2 \leqslant w_{12} \leqslant p_1 + p_2 - 2p(000),$$

$$0 \leqslant p(010) \leqslant 1$$
$$\Leftrightarrow p_1 + p_3 - 2 \leqslant w_{13} \leqslant p_1 + p_3 - 2p(000),$$

$$0 \leqslant p(100) \leqslant 1$$
$$\Leftrightarrow p_2 + p_3 - 2 \leqslant w_{23} \leqslant p_2 + p_3 - 2p(000),$$

$$0 \leqslant p(011) \leqslant 1$$
$$\Leftrightarrow p_2 + p_3 - 2p(000) \leqslant w_{12} + w_{13} \leqslant 2 - 2p(000)$$
$$+ p_2 + p_3,$$

$$0 \leqslant p(101) \leqslant 1$$
$$\Leftrightarrow p_1 + p_3 - 2p(000) \leqslant w_{12} + w_{23} \leqslant 2 - 2p(000)$$
$$+ p_1 + p_3,$$

$$0 \leqslant p(110) \leqslant 1$$
$$\Leftrightarrow p_1 + p_2 - 2p(000) \leqslant w_{13} + w_{23} \leqslant 2 - 2p(000)$$
$$+ p_1 + p_2,$$

$$0 \leqslant p(111) \leqslant 1$$
$$\Leftrightarrow -2p(000) \leqslant w_{12} + w_{23} + w_{13} \leqslant 2 - 2p(000).$$

Combining the inequalities to eliminate $p(000)$, and using the fact that $0 \leqslant p(000) \leqslant 1$:

$$0 \leqslant p(000) \leqslant 1, 0 \leqslant p(111) \leqslant 1$$
$$\Rightarrow -2p(000) \leqslant 0 \leqslant w_{12} + w_{23} + w_{13} \leqslant 2 - 2p(000) \leqslant 2,$$

$$0 \leqslant p(010) \leqslant 1, 0 \leqslant p(101) \leqslant 1$$
$$\Rightarrow 0 \leqslant w_{12} + w_{23} - w_{13} \leqslant 2 \leqslant 4 - 2p(000),$$

$$0 \leqslant p(110) \leqslant 1, 0 \leqslant p(001) \leqslant 1$$
$$\Rightarrow 0 \leqslant w_{23} + w_{13} - w_{12} \leqslant 2 \leqslant 4 - 2p(000),$$

$$0 \leqslant p(011) \leqslant 1, 0 \leqslant p(100) \leqslant 1$$
$$\Rightarrow 0 \leqslant w_{12} + w_{13} - w_{23} \leqslant 2 \leqslant 4 - 2p(000).$$

Of these, the KS inequalities, which are not trivially true by normalization and positivity, are the following:

$$R_3 \equiv w_{12} + w_{23} + w_{13} \leqslant 2, \tag{B1}$$

$$R_0 \equiv w_{12} - w_{23} - w_{13} \leqslant 0, \tag{B2}$$

$$R_1 \equiv w_{23} - w_{12} - w_{13} \leqslant 0, \tag{B3}$$

$$R_2 \equiv w_{13} - w_{12} - w_{23} \leqslant 0. \tag{B4}$$

Note that $p(000) \leqslant p_1, p_2, p_3$, since $p_1 = p(000) + p(001) + p(010) + p(011)$, etc. As long as the KS inequalities are

satisfied, one can define a joint probability distribution by choosing a suitable $p(000) \leqslant \min\{p_1, p_2, p_3\}$.

To summarize, there are following constraints on the six parameters, $\{w_{12}, w_{23}, w_{13}, p_1, p_2, p_3\}$, characterizing the polytope of KS-noncontextual correlations:

$$0 \leqslant p_1, p_2, p_3, w_{12}, w_{23}, w_{13} \leqslant 1, \tag{B5}$$

$$|p_1 - p_2| \leqslant w_{12} \leqslant \min\{p_1 + p_2, 2 - p_1 - p_2\}, \tag{B6}$$

$$|p_2 - p_3| \leqslant w_{23} \leqslant \min\{p_2 + p_3, 2 - p_2 - p_3\}, \tag{B7}$$

$$|p_1 - p_3| \leqslant w_{13} \leqslant \min\{p_1 + p_3, 2 - p_1 - p_3\}, \tag{B8}$$

$$w_{12} + w_{23} + w_{13} \leqslant 2, \tag{B9}$$

$$w_{12} - w_{23} - w_{13} \leqslant 0, \tag{B10}$$

$$w_{23} - w_{12} - w_{13} \leqslant 0, \tag{B11}$$

$$w_{13} - w_{12} - w_{23} \leqslant 0. \tag{B12}$$

2. NC inequalities

In deriving the NC inequalities, I closely follow the derivation of the LSW inequality in Ref. [10]. For a more detailed explication of the principles underlying this derivation, the reader may consult Ref. [13]. The assumptions used are: measurement noncontextuality and preparation noncontextuality on account of the fact that in operational quantum theory, preparation noncontextuality implies outcome determinism for sharp measurements [2].

Define

$$R_3(\lambda) \equiv w_{12}(\lambda) + w_{23}(\lambda) + w_{13}(\lambda), \tag{B13}$$

$$R_0(\lambda) \equiv w_{12}(\lambda) - w_{23}(\lambda) - w_{13}(\lambda), \tag{B14}$$

$$R_1(\lambda) \equiv w_{23}(\lambda) - w_{12}(\lambda) - w_{13}(\lambda), \tag{B15}$$

$$R_2(\lambda) \equiv w_{13}(\lambda) - w_{12}(\lambda) - w_{23}(\lambda), \tag{B16}$$

where $w_{ij}(\lambda) \equiv \xi(X_i \neq X_j | M_{ij}; \lambda)$, for all $(ij) \in \{(12), (23), (13)\}$. Note that for any given preparation, the ontological model associates a positive density $\mu(\lambda|P) \geqslant 0$, where $\int d\lambda \mu(\lambda|P) = 1$, and $p(X|M, P) = \int d\lambda \mu(\lambda|P)\xi(X|M; \lambda)$, where $\xi(X|M; \lambda) \in [0, 1]$ is the response function of outcome X when measurement M is performed and the system's ontic state is λ. Therefore,

$$R_3 = \int d\lambda \mu(\lambda|P) R_3(\lambda) \leqslant \max_\lambda R_3(\lambda), \tag{B17}$$

$$R_0 = \int d\lambda \mu(\lambda|P) R_0(\lambda) \leqslant \max_\lambda R_0(\lambda), \tag{B18}$$

$$R_1 = \int d\lambda \mu(\lambda|P) R_1(\lambda) \leqslant \max_\lambda R_1(\lambda), \tag{B19}$$

$$R_2 = \int d\lambda \mu(\lambda|P) R_2(\lambda) \leqslant \max_\lambda R_2(\lambda). \tag{B20}$$

To maximize $R_3(\lambda)$ in this noncontextual model, one needs to maximize each anticorrelation term $w_{12}(\lambda), w_{23}(\lambda), w_{13}(\lambda)$. To maximize $R_0(\lambda)$, maximize $w_{12}(\lambda)$ and minimize $w_{23}(\lambda), w_{13}(\lambda)$. Similarly, to maximize $R_1(\lambda)$, maximize $w_{23}(\lambda)$ and minimize $w_{12}(\lambda), w_{13}(\lambda)$, and to maximize $R_2(\lambda)$, maximize $w_{13}(\lambda)$ and minimize $w_{12}(\lambda), w_{23}(\lambda)$.

The single measurement response functions are given by

$$\xi(X_i|M_i; \lambda) = \eta \delta_{X_i, X_i(\lambda)} + (1 - \eta)\left(\tfrac{1}{2}\delta_{X_i, 0} + \tfrac{1}{2}\delta_{X_i, 1}\right), \tag{B21}$$

$i \in \{1, 2, 3\}$, in keeping with the assumption of outcome determinism for projectors but not so for nonprojective positive operators [13]. The general form the pairwise response function for measurements $\{M_i, M_j\}$ is given by

$$\begin{aligned}
\xi(X_i, X_j | M_{ij}; \lambda) &= \alpha \delta_{X_i, X_i(\lambda)} \delta_{X_j, X_j(\lambda)} \\
&\quad + \beta \delta_{X_i, X_i(\lambda)}\left(\tfrac{1}{2}\delta_{X_j, 0} + \tfrac{1}{2}\delta_{X_j, 1}\right) \\
&\quad + \gamma \left(\tfrac{1}{2}\delta_{X_i, 0} + \tfrac{1}{2}\delta_{X_i, 1}\right)\delta_{X_j, X_j(\lambda)} \\
&\quad + \delta \left(\tfrac{1}{2}\delta_{X_i, 0}\delta_{X_j, 0} + \tfrac{1}{2}\delta_{X_i, 1}\delta_{X_j, 1}\right) \\
&\quad + \epsilon \left(\tfrac{1}{2}\delta_{X_i, 0}\delta_{X_j, 1} + \tfrac{1}{2}\delta_{X_i, 1}\delta_{X_j, 0}\right). \tag{B22}
\end{aligned}$$

The marginals are

$$\begin{aligned}
\xi(X_i | M_{ij}; \lambda) &= (\alpha + \beta)\delta_{X_i, X_i(\lambda)} \\
&\quad + (\gamma + \delta + \epsilon)\left(\tfrac{1}{2}\delta_{X_i, 0} + \tfrac{1}{2}\delta_{X_i, 1}\right) \tag{B23}
\end{aligned}$$

and

$$\begin{aligned}
\xi(X_j | M_{ij}; \lambda) &= (\alpha + \gamma)\delta_{X_j, X_j(\lambda)} \\
&\quad + (\beta + \delta + \epsilon)\left(\tfrac{1}{2}\delta_{X_j, 0} + \tfrac{1}{2}\delta_{X_j, 1}\right), \tag{B24}
\end{aligned}$$

so that the following must hold on account of $\xi(X_i|M_{ij}; \lambda) = \xi(X_i|M_i; \lambda)$ and $\xi(X_j|M_{ij}; \lambda) = \xi(X_j|M_j; \lambda)$:

$$\alpha + \beta = \alpha + \gamma = \eta, \tag{B25}$$

$$\gamma + \delta + \epsilon = \beta + \delta + \epsilon = 1 - \eta. \tag{B26}$$

To maximize anticorrelation $w_{ij}(\lambda)$: the β and γ terms yield correlation as often as anticorrelation, so $\beta = \gamma = 0$. The δ term always yields correlation, so $\delta = 0$. Only α and ϵ terms allow for more anticorrelation than correlation. This means $\alpha = \eta$ and $\epsilon = 1 - \eta$. The pairwise response function maximizing anticorrelation $w_{ij}(\lambda)$ is given by

$$\begin{aligned}
\xi(X_i X_j | M_{ij}; \lambda) &= \eta \delta_{X_i, X_i(\lambda)} \delta_{X_j, X_j(\lambda)} \\
&\quad + (1 - \eta)\left(\tfrac{1}{2}\delta_{X_i, 0}\delta_{X_j, 1} + \tfrac{1}{2}\delta_{X_i, 1}\delta_{X_j, 0}\right).
\end{aligned}$$

This maximizing response function constrains the anticorrelation probability as

$$1 - \eta \leqslant w_{ij}(\lambda) \leqslant 1. \tag{B27}$$

To minimize anticorrelation $w_{ij}(\lambda)$: the β and γ terms yield correlation as often as anticorrelation, so $\beta = \gamma = 0$. The ϵ term always yields anticorrelation, so $\epsilon = 0$. Only α and δ terms allow for more correlation than anticorrelation. This means $\alpha = \eta$ and $\delta = 1 - \eta$. The pairwise response function minimizing anticorrelation $w_{ij}(\lambda)$ is given by

$$\begin{aligned}
\xi(X_i X_j | M_{ij}; \lambda) &= \eta \delta_{X_i, X_i(\lambda)} \delta_{X_j, X_j(\lambda)} \\
&\quad + (1 - \eta)\left(\tfrac{1}{2}\delta_{X_i, 0}\delta_{X_j, 0} + \tfrac{1}{2}\delta_{X_i, 1}\delta_{X_j, 1}\right).
\end{aligned}$$

This minimizing response function constrains the anticorrelation probability as

$$0 \leqslant w_{ij}(\lambda) \leqslant \eta. \tag{B28}$$

$R_3(\lambda)$ is maximized by considering the response function maximizing anticorrelation for each of $w_{ij}(\lambda)$, and noting that of the eight possible assignments $\lambda \to (X_1(\lambda), X_2(\lambda), X_3(\lambda)) \in \{0,1\}^3$, the assignments maximizing $R_3(\lambda)$ are $\{(001),(010),(011),(100),(101),(110)\}$, each of which has two anticorrelated pairs and a third correlated pair such that the anticorrelation probability becomes $\max_\lambda R_3(\lambda) = 2\eta + 3(1 - \eta) = 3 - \eta$, and therefore

$$R_3 \leqslant 3 - \eta, \tag{B29}$$

the LSW inequality.

$R_0(\lambda)$ is maximized by considering the response function maximizing anticorrelation for $w_{12}(\lambda)$ and response functions minimizing anticorrelation for $w_{23}(\lambda)$ and $w_{13}(\lambda)$. Noting that of the eight possible assignments $\lambda \to (X_1(\lambda), X_2(\lambda), X_3(\lambda)) \in \{0,1\}^3$, the assignments maximizing $R_0(\lambda)$ are $\{(010),(011),(100),(101)\}$: for $\{(010),(101)\}$, $w_{12}(\lambda) = 1$, $w_{23}(\lambda) = \eta$, and $w_{13}(\lambda) = 0$, and for $\{(011),(100)\}$, $w_{12}(\lambda) = 1$, $w_{23}(\lambda) = 0$, and $w_{13}(\lambda) = \eta$, so that $\max_\lambda R_0(\lambda) = 1 - \eta$, and therefore

$$R_0 \leqslant 1 - \eta. \tag{B30}$$

Similarly, the NC inequalities for R_1 and R_2 follow: $R_1 \leqslant 1 - \eta$ and $R_2 \leqslant 1 - \eta$.

---

[1] S. Kochen and E. Specker, The problem of hidden variables in quantum mechanics, Indiana University Mathematics J. **17**, 59 (1968).

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

[3] J. S. Bell, On the Einstein Podolsky Rosen paradox, Physics **1**, 195 (1964); Reprinted in *Speakable and Unspeakable in Quantum Mechanics* (Cambridge University Press, New York, 1987), Chap. 2.

[4] J. S. Bell, The theory of local beables, Epistemolog. Lett. **9** (1976); Reproduced in *John S Bell on the Foundations of Quantum Mechanics*, edited by M. Bell, K. Gottfried, and M. Veltman (World Scientific, Singapore, 2001), p. 50.

[5] H. M. Wiseman, The two Bell's theorems of John Bell, J. Phys. A: Math. Theor. **47**, 424001 (2014).

[6] A. Fine, Hidden variables, joint probability, and the Bell inequalities, Phys. Rev. Lett. **48**, 291 (1982).

[7] A. Fine, Joint distributions, quantum correlations, and commuting observables, J. Math. Phys. **23**, 1306 (1982).

[8] E. P. Specker, Dialectica **14**, 239 (1960); English translation: M. P. Seevinck, The logic of non-simultaneously decidable propositions, arXiv:1103.4537.

[9] J. S. Bell, On the problem of hidden variables in quantum mechanics, Rev. Mod. Phys. **38**, 447 (1966).

[10] Y. C. Liang, R. W. Spekkens, and H. M. Wiseman, Specker's parable of the overprotective seer: A road to contextuality, nonlocality and complementarity, Phys. Rep. **506**, 1 (2011).

[11] R. Kunjwal and S. Ghosh, Minimal state-dependent proof of measurement contextuality for a qubit, Phys. Rev. A **89**, 042118 (2014).

[12] J. Barrett, L. Hardy, and A. Kent, No signaling and quantum key distribution, Phys. Rev. Lett. **95**, 010503 (2005).

[13] R. W. Spekkens, The status of determinism in proofs of the impossibility of a noncontextual model of quantum theory, Found. Phys. **44**, 1125 (2014).

[14] S. Abramsky and A. Brandenburger, The sheaf-theoretic structure of non-locality and contextuality, New J. Phys. **13**, 113036 (2011).

[15] M. Araújo, M. T. Quintino, C. Budroni, M. T. Cunha, and A. Cabello, All noncontextuality inequalities for the *n*-cycle scenario, Phys. Rev. A **88**, 022118 (2013).

[16] R. Chaves and T. Fritz, Entropic approach to local realism and noncontextuality, Phys. Rev. A **85**, 032113 (2012).

[17] R. Kunjwal and R. W. Spekkens, Noncontextuality without determinism and admissible (in)compatibility relations: revisiting Specker's parable, see also Perimenter institute seminar (2014), PIRSA:14010102, http://pirsa.org/14010102.