

# Classical Causal Models for Bell and Kochen-Specker Inequality Violations Require Fine-Tuning

Eric G. Cavalcanti\*

Centre for Quantum Computation and Communication Technology (Australian Research Council),  
Centre for Quantum Dynamics, Griffith University, Brisbane, Queensland 4111, Australia

 (Received 6 October 2017; revised manuscript received 19 January 2018; published 13 April 2018)

Nonlocality and contextuality are at the root of conceptual puzzles in quantum mechanics, and they are key resources for quantum advantage in information-processing tasks. Bell nonlocality is best understood as the incompatibility between quantum correlations and the classical theory of causality, applied to relativistic causal structure. Contextuality, on the other hand, is on a more controversial foundation. In this work, I provide a common conceptual ground between nonlocality and contextuality as violations of classical causality. First, I show that Bell inequalities can be derived solely from the assumptions of no signaling and no fine-tuning of the causal model. This removes two extra assumptions from a recent result from Wood and Spekkens and, remarkably, does not require any assumption related to independence of measurement settings—unlike all other derivations of Bell inequalities. I then introduce a formalism to represent contextuality scenarios within causal models and show that all classical causal models for violations of a Kochen-Specker inequality require fine-tuning. Thus, the quantum violation of classical causality goes beyond the case of spacelike-separated systems and already manifests in scenarios involving single systems.

DOI: 10.1103/PhysRevX.8.021018

Subject Areas: Quantum Physics, Quantum Information

## I. INTRODUCTION

Quantum contextuality, the phenomenon uncovered by Kochen and Specker (KS) [1], is at the core of the quantum departure from classicality and has recently been identified as a candidate for the resource behind the power of quantum computation [2]. Much controversy still exists, however, on what exactly contextuality is, with different formalisms giving different definitions of the phenomenon [3–6]. For example, derivations following the work of Kochen-Specker require an assumption of outcome determinism, the validity of which in experimentally relevant situations has been criticized [3,7]. Indeed, it has been argued that it is not possible to experimentally test contextuality without extra assumptions [8].

Bell nonlocality [9] rests on comparatively solid foundations. It is best understood as the incompatibility between quantum correlations and causal constraints [10]. A modern approach is to capture these constraints within the framework of causal networks [11], where causal structure is represented as a directed acyclic graph (DAG) (Fig. 1). Assuming a causal graph motivated by relativity (Fig. 2),

we can derive constraints on observable probability distributions—the Bell inequalities. A violation of a Bell inequality thus implies either a violation of relativistic causality or of one or more of the assumptions underlying this framework for causality, such as Reichenbach’s principle of common cause [12,13]. This second alternative has motivated a program to extend the classical causal formalism to a framework of quantum causal models [13–18], opening the exciting prospect of a coherent understanding of the nature of causality in a quantum world and a resolution of at least part of the puzzle of Bell’s theorem [19].

Contextuality, on the other hand, is *a priori* unrelated to causality: It is not necessary that measurements are spacelike separated or that they involve separate subsystems at all. Thus, it is not clear how a theory of quantum causality could help with contextuality. Here, I bridge that gap and show that all classical causal models that reproduce the violation of Bell and KS inequalities violate a core principle of the causal-models framework: no fine-tuning. This unifies Bell nonlocality and KS contextuality as

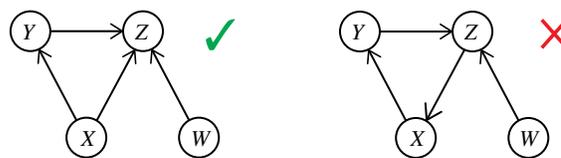


FIG. 1. In a DAG (left diagram), nodes represent random variables, and directed edges represent causal links. Closed cycles (right diagram) are excluded.

\*e.cavalcanti@griffith.edu.au

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.

violations of classical causality and opens a new direction for the study of contextuality as a resource.

In an influential work [20], Wood and Spekkens showed that every classical causal model that can reproduce Bell inequality violations requires causal connections not observed in the phenomena, such as faster-than-light causation. Only with special, finely tuned parameters can a causal model “hide” those connections from observers. This provides a novel way of looking at Bell inequalities—not as implications of relativistic causal structure but as implications of classical causal principles for *any* causal structure. That result, however, applied only to phenomena satisfying two extra assumptions, which makes it inapplicable to contextuality scenarios. Here, I prove a more general result without those extra conditions and show how it can be extended to the case of contextuality.

This work is organized as follows. In Sec. II, I review the framework of causal models and show how it can be used to derive Bell inequalities by assuming free choice and relativistic causality. In Sec. III, I introduce a formalism for describing contextuality scenarios within classical causal models and present the main results. In Sec. IV, I illustrate the need for fine-tuning in causal models for violations of Kochen-Specker inequalities with an example. In Sec. V, I discuss the relation between the present result and the previous work of Ref. [20] and how it provides a novel interpretation for quantum advantage in biased nonlocal games [21]. I conclude in Sec. VI with a summary of the results and a discussion of its relevance, drawbacks, and directions for further research.

## II. BACKGROUND

### A. Causal models

A modern framework for causation and its role in explaining correlations can be found in the theory of causal networks [11]. With extensive applicability from statistics to epidemiology, economics, and artificial intelligence, it has been developed as a tool to connect causal inferences and probabilistic observations. In such a model, causal structure is represented as a graph  $\mathcal{G}$ , with variables as nodes and direct causal links as directed edges (arrows) between nodes. To avoid the potential for paradoxical causal loops, closed cycles are forbidden, and the resulting structure is a DAG (Fig. 1). The relations between nodes in a DAG  $\mathcal{G}$  can be expressed in an intuitive genealogical terminology: Nodes pointing to a given node  $X$  (the direct causes of  $X$ ) are called the parents of  $X$ , denoted as  $\text{Pa}(X)$ ; the ancestors of  $X$ ,  $\text{An}(X)$ , are all nodes from which there is a directed path to  $X$  (i.e., all variables in the causal past of  $X$ ); the descendants of  $X$ ,  $\text{De}(X)$ , are all nodes for which  $X$  is an ancestor (i.e., all variables in the causal future of  $X$ ); the set of nondescendants of  $X$  is denoted by  $\text{Nd}(X)$ .

The purpose of a DAG is to encode the conditional independences associated with any probability distribution

compatible with the causal structure, through the *causal Markov condition*: In any probability distribution  $P$  that is *compatible* with a graph  $\mathcal{G}$ , a variable  $X$  is independent of all of its nondescendants, conditional on its parents. In other words,  $P(X|\text{Nd}(X), \text{Pa}(X)) = P(X|\text{Pa}(X))$ , which we denote as  $(X \perp\!\!\!\perp \text{Nd}(X)|\text{Pa}(X))$ . The causal Markov condition is equivalent to the requirement that any distribution over the variables  $X_1, \dots, X_n$  compatible with the graph  $\mathcal{G}$  factorizes as

$$P(X_1, \dots, X_n) = \prod_j P(X_j|\text{Pa}(X_j)). \quad (1)$$

Those conditional independences can be obtained from the graph through a rule called *d-separation* [11]. Two sets of variables  $X$  and  $Y$  are *d-separated* given a set of variables  $Z$  [denoted  $(X \perp\!\!\!\perp Y|Z)_d$ ] if and only if  $Z$  “blocks” all paths  $p$  from  $X$  to  $Y$ . A path  $p$  is blocked by  $Z$  if and only if (i) it contains a chain  $A \rightarrow B \rightarrow C$  or a fork  $A \leftarrow B \rightarrow C$  such that the middle node  $B$  is in  $Z$ , or (ii) it contains an inverted fork (head-to-head)  $A \rightarrow B \leftarrow C$  such that the node  $B$  is not in  $Z$ , and there is no directed path from  $B$  to any member of  $Z$ .

The *d-separation* is a sound and complete criterion for conditional independence: If in a DAG  $\mathcal{G}$  two variables  $X$  and  $Y$  are *d-separated* given  $Z$ ,  $(X \perp\!\!\!\perp Y|Z)_d$ , then they are conditionally independent given  $Z$ ,  $(X \perp\!\!\!\perp Y|Z)$ , in all distributions compatible with  $\mathcal{G}$ ; if for all distributions compatible with  $\mathcal{G}$ , the conditional independence  $(X \perp\!\!\!\perp Y|Z)$  holds, then  $\mathcal{G}$  satisfies  $(X \perp\!\!\!\perp Y|Z)_d$ .

### B. Causal models and Bell’s theorem

As an example of the application of this framework, I review how it can be used to derive Bell’s theorem. Consider the correlations between measurements performed by two agents, Alice and Bob. Alice’s choice of measurement is represented by a variable  $X$ , and Bob’s by a variable  $Y$ . Their respective outcomes are represented by  $A$  and  $B$ . Their measurements are assumed to be performed within spacelike-separated regions, so no relativistic causal connection can exist between the variables in Alice’s lab and those in Bob’s lab. They may however be correlated because of variables in their common causal past, the set of which is denoted by  $\Lambda$ . The assumption that Alice and Bob can make “free choices” is translated as the requirement that  $X$  and  $Y$  are exogenous variables: They have no relevant causes. This scenario is represented in the graphical notation as in Fig. 2.

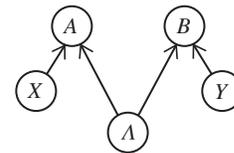


FIG. 2. DAG representing the causal structure of a Bell scenario.

The causal Markov condition then implies that  $P(AB\Lambda|XY) = P(A|X\Lambda)P(B|Y\Lambda)P(\Lambda)$ . Averaging over  $\Lambda$ , we obtain the factorizability condition of a local hidden variable model:

$$P(AB|XY) = \sum_{\Lambda} P(\Lambda)P(A|X\Lambda)P(B|Y\Lambda). \quad (2)$$

As is well known, this leads to the Bell inequalities, which can be violated by quantum correlations [22]. Therefore, assuming relativistic causal structure and free choices, quantum correlations cannot be reproduced by the classical framework of causality.

Note that the assumption of free choice is not strictly necessary: A weaker but sufficient condition is simply “ $\Lambda$ -independence”; i.e., the measurement choices are independent of any latent variables  $\Lambda$  that are causally connected with the systems, ( $\Lambda \perp\!\!\!\perp XY$ ). This would still be compatible with a causal graph where there is a common cause between the measurement choices  $X$  and  $Y$ . But without one of these assumptions, it would be possible for a conspiratorial “superdeterministic” theory to reproduce the quantum correlations. Remarkably, in the main result of this paper, *neither* of these assumptions is needed. Superdeterministic theories, for example, are ruled out because they violate the no-fine-tuning rule.

### III. CONTEXTUALITY SCENARIOS AND CAUSAL MODELS

#### A. Basic definitions

Traditionally, locality and noncontextuality have been spelled out in terms of ontological models, but here I translate those concepts into the language of causal models. Indeed, we can think of ontological models as causal models in disguise, which is a useful perspective as it allows one to more clearly identify implicit classical causal assumptions that may be revised in light of quantum causal models. The formalism used here is most analogous to that of Abramsky and Brandenburger [4], although it is expressed in the language of causal models and uses simplified terminology (for example, without reference to sheaf theory).

A measurement scenario is specified by the following: (i) a set  $\mathcal{M}$  of measurements; (ii) for each measurement  $m \in \mathcal{M}$ , a set of possible outcomes  $\mathcal{O}_m$ ; and (iii) a compatibility structure  $\mathcal{C}$  on  $\mathcal{M}$ —a family of subsets of  $\mathcal{M}$ —specifying joint measurability. Two measurements  $m_1, m_2 \in \mathcal{M}$  are said to be jointly measurable or compatible, or to be part of a context, iff  $\{m_1, m_2\} \in \mathcal{C}$ , and likewise for sets of more than two measurements [23]. Without loss of generality, we can enlarge each  $\mathcal{O}_m$  so that all measurements have the same number of outcomes, and we can label them so that all outcome sets are equal and denoted simply by  $\mathcal{O}$ .

Consider now an individual test within a measurement scenario, where a set of  $n$  random variables,  $X_1, \dots, X_n$ , specifies  $n$  measurements to be performed upon the system. A compatibility scenario, or contextuality scenario, is one in which, for any given run,  $x_1, \dots, x_n$  are jointly measurable. In other words, let  $X_i = x_i \in \mathcal{M}$  denote the values of those variables in a particular run. Then,  $\{x_1, \dots, x_n\} \in \mathcal{C}$ . This could be done, for example, through a further random variable  $C$  that selects a context, from which  $X_1, \dots, X_n$  are determined, but I make no assumption about the process of selection or the order in which the measurements are performed. A special class of contextuality scenarios is that in which  $\mathcal{M}$  can be decomposed into  $k$  subsets  $\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_k$  such that all contexts  $c \in \mathcal{C}$  have at most one element from each subset. These are called ( $k$ -partite) Bell-nonlocality scenarios.

From here on, I consider scenarios containing only *pairs* of compatible measurements. This does not necessarily mean that there are no sets of three or more compatible measurements that can be performed on the system but only that I am restricting  $\mathcal{M}$  so that it only contains pairs of them. These will be called binary contextuality scenarios, a special case of which are bipartite Bell scenarios. The measurements will be chosen through random variables  $X$  and  $Y$ , with outcomes respectively recorded by random variables  $A$  and  $B$ .

A phenomenon  $\mathcal{P}$  for such a scenario is specified by a probability distribution  $\mathcal{P}(ABXY)$  for the observable variables. Note that no causal assumption has been made at this stage. I now define what I mean by a (classical) causal model for a phenomenon.

**Definition 1** (Causal model). A (classical) causal model  $\Gamma$  for a phenomenon  $\mathcal{P}$  consists of a (possibly empty) set of latent variables  $\Lambda$ , a DAG  $\mathcal{G}$  with nodes  $\{A, B, X, Y, \Lambda\}$ , and a probability distribution  $P(ABXY\Lambda)$  compatible with  $\mathcal{G}$ , such that  $\mathcal{P}(ABXY) = \sum_{\Lambda} P(ABXY\Lambda)$ .

A special class of phenomena are those where the probability for the outcome of one measurement does not depend on which measurement it may be performed together with, i.e., those that satisfy the property of no disturbance.

**Definition 2** (No disturbance). A phenomenon is said to satisfy *no-disturbance* iff  $\mathcal{P}(A|XY) = \mathcal{P}(A|X)$  and  $\mathcal{P}(B|XY) = \mathcal{P}(B|Y)$  for all values of the variables  $A, B, X, Y$  for which those conditionals are defined.

In the causal-model notation, the no-disturbance conditions are denoted by  $(A \perp\!\!\!\perp Y|X)$  and  $(B \perp\!\!\!\perp X|Y)$ . Compatibility scenarios naturally satisfy those conditions since  $X$  and  $Y$  can only take joint values as pairs of compatible measurements and since the no-disturbance condition is implicit in the very meaning of compatibility [4]. In Bell scenarios, this assumption is called no signaling, and where the measurements  $X$  and  $Y$  are performed in spacelike-separated regions, it is justified by relativity.

Whereas no-disturbance and no-signaling are purely properties of *phenomena*, the subsequent definitions are

about properties of *causal models* for phenomena. I say that a phenomenon violates a property when no causal model for the phenomenon satisfies that property.

Bell locality and KS noncontextuality are equivalent to the existence of a factorizable hidden variable model, which, in the language of causal models, translates to the following:

**Definition 3** (Factorizability). A causal model is said to satisfy factorizability iff  $\forall A, B, X, Y, P(AB|XY) = \sum_{\Lambda} P(\Lambda)P(A|X\Lambda)P(B|Y\Lambda)$ .

The original Kochen-Specker theorem assumed the existence of a deterministic noncontextual model. By Fine's theorem [25], this is equivalent to the existence of a factorizable model, and to the existence of a joint probability distribution for the outcomes of all measurements that returns the observable correlations as marginals.

**Definition 4** (KS noncontextuality). A causal model for a contextuality scenario is said to satisfy KS noncontextuality iff it is factorizable.

From factorizability, one can derive, for each contextuality scenario, inequalities that bound the set of KS-noncontextual phenomena, as facets of a convex polytope [26,27]. These are the KS inequalities [28], which reduce to Bell inequalities in Bell scenarios.

Despite this formal equivalence, much controversy remains regarding the justification for factorizability in contextuality scenarios. While in Bell scenarios it is implied by Bell's notion of local causality (for a detailed review, see Ref. [10]), this is not the case in contextuality scenarios, where measurements are not spacelike separated. To derive factorizability in this case, one requires, apart from the assumption of measurement noncontextuality, an extra assumption of outcome determinism [7], calling into question the implication of violations of Kochen-Specker inequalities. However, as we will see below, factorizability is implied by another fundamental principle of the causal-models framework, the principle of no-fine-tuning, or faithfulness.

**Definition 5** (No fine-tuning). A causal model  $\Gamma$  is said to satisfy no fine-tuning, or be *faithful*, relative to a phenomenon  $\mathcal{P}$  iff every conditional independence  $(C \perp\!\!\!\perp D|E)$  in  $\mathcal{P}$  corresponds to a  $d$ -separation  $(C \perp\!\!\!\perp D|E)_d$  in the causal graph  $\mathcal{G}$  of  $\Gamma$ .

For example, suppose a phenomenon satisfies the no-disturbance condition  $(A \perp\!\!\!\perp Y|X)$ , but its causal structure contains a direct link from  $Y$  to  $A$  and thus lacks the  $d$ -separation  $(A \perp\!\!\!\perp Y|X)_d$ . This is only possible if some of the parameters of the model take special values that hide the influence of  $Y$  on  $A$ . An example is Bohmian mechanics, where the probability distributions for the hidden variables are constrained by the “quantum equilibrium” condition [34]. A faithful causal model, on the other hand, has no such hidden causal connections.

Therefore, the no-fine-tuning condition can be seen as an instance of Occam's razor [11], or of a methodological

principle that also motivates Spekkens' notion of generalized contextuality [3]: Leibniz's principle of the identity of indiscernibles. This principle states that one should not postulate differences in the ontological (i.e., causal-model) description of a phenomenon where none exists at the operational level. Here, however, I show that this principle leads to KS noncontextuality, implying that the two competing notions of noncontextuality have more in common than has been previously realized. Whether Spekkens' notion can also be directly derived from faithfulness is an interesting question left for future work.

## B. Main result

**Theorem 1.** Every faithful causal model for a no-disturbance phenomenon is factorizable.

As shown in detail in the Appendix, the proof of Theorem 1 proceeds by showing that all DAGs that do not require fine-tuning to explain the no-disturbance conditions lead to factorizability. We first note that the no-disturbance conditions, together with the assumption of no fine-tuning, imply that every DAG  $\mathcal{G}$  for a no-disturbance phenomenon  $\mathcal{P}$  must satisfy the  $d$ -separation conditions  $(A \perp\!\!\!\perp Y|X)_d$  and  $(B \perp\!\!\!\perp X|Y)_d$ . We thus proceed by excluding every DAG that does not satisfy these conditions, and I show that all remaining DAGs imply factorizability. As an immediate corollary:

**Corollary 1.** No fine-tuning and no-disturbance imply KS noncontextuality.

Another corollary is a stronger version of the result of Ref. [20], without the extra assumptions of marginal-setting independence and local-setting dependence:

**Corollary 2.** No fine-tuning and no-signaling imply Bell locality.

It is instructive to state Corollary 1 in a contrapositive form:

**Corollary 3.** Every classical causal model that reproduces the violation of a KS inequality in a no-disturbance phenomenon requires fine-tuning.

This result implies that the quantum violation of classical causality, long recognized in the case of spacelike-separated entangled quantum systems, also manifests in the case of single systems.

## IV. EXAMPLES

To illustrate the need for fine-tuning in causal explanations of KS-inequality violations, let us consider as a simple example the three-observable scenario introduced by Liang, Wiseman, and Spekkens (LSW) in Ref. [24]. The measurement scenario consists of a set of three measurements  $\mathcal{M} = \{m_1, m_2, m_3\}$  with binary outcomes  $\mathcal{O} = \{0, 1\}$  and compatibility structure  $\mathcal{C} = \{\{m_1, m_2\}, \{m_1, m_3\}, \{m_2, m_3\}\}$ —that is, they are pairwise compatible but not triplewise compatible. In any given experimental test,  $X$  and  $Y$  can take any pair of values from  $\mathcal{C}$  and  $A$ , and  $B$  can take values

from  $\mathcal{O} = \{0, 1\}$ . Consider a phenomenon  $\mathcal{P}(ABXY)$  that satisfies the no-disturbance relations  $\mathcal{P}(A|XY) = \mathcal{P}(A|X)$  and  $\mathcal{P}(B|XY) = \mathcal{P}(B|Y)$ . Thus, from Theorem 1, any faithful causal model for this phenomenon must satisfy KS noncontextuality.

In this scenario, KS noncontextuality implies the LSW inequality [24]:

$$\sum_{\{m_i, m_j\} \in \mathcal{C}} \frac{1}{3} \mathcal{P}(A \neq B | m_i, m_j) \leq \frac{2}{3}. \quad (3)$$

To see this, recall that the existence of a KS-noncontextual model is equivalent to the existence of a joint probability distribution for the outcomes of all measurements. For each extreme point of this distribution, at most two of the three pairs of measurements can be anticorrelated, so the average probability of anticorrelation cannot exceed  $2/3$ . [Note that since  $(X, Y) = (m_1, m_2)$  and  $(X, Y) = (m_2, m_1)$  correspond to the same pair of measurements being performed,  $\mathcal{P}(A = k, B = l | X = m_i, Y = m_j) = \mathcal{P}(A = l, B = k | X = m_j, Y = m_i)$ .]

The authors of Ref. [24] illustrate a hypothetical maximal violation of Eq. (3) as a realization of the ‘‘parable of the overprotective seer’’ (OS): Wishing to ward off unworthy suitors for his beloved daughter, a seer from ancient Assyria proposed to each of them the following task. They were taken to a table upon which sat three boxes and asked to open two of the boxes. If both boxes contained a gem and both boxes did not contain a gem, they would win the task and would be allowed to marry the seer’s daughter. As it turned out, every suitor always randomly found a gem in one box and none in the other—a seemingly paradoxical situation.

An example of a causal model that realizes the OS correlations consists of the causal graph in Fig. 3 and the following causal parameters. A uniformly distributed latent variable  $\Lambda$  determines, for each box, whether or not it contains a gem. Let  $X$  denote the first box opened by the suitor, so  $A(X, \Lambda) \in \{0, 1\}$  for any value of  $X$ , with  $A = 1$  representing the presence of a gem. Depending on whether or not the first box contained a gem, a gem is added or removed from the second box  $Y$  to ensure that  $B(X, Y, \Lambda) = A(X, \Lambda) \oplus 1$  for all  $\{X, Y\} \in \mathcal{C}$ . To see the need for fine-tuning, note that if one were able to prepare any distribution for the hidden variables, the no-disturbance condition would be violated. For example, suppose one could prepare  $\Lambda = \Lambda_0$  such that  $A(X = m_i, \Lambda_0) = \delta_{i,1}$ . If

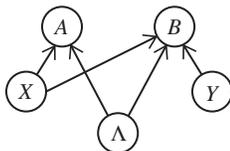


FIG. 3. Example of a causal structure that can reproduce violations of KS inequalities with fine-tuning.

$Y = m_3$  is measured, its outcome will be  $B = 0$  if  $X = m_1$  and  $B = 1$  if  $X = m_2$ . Of course, the need for fine-tuning would manifest in different ways for other causal models, but Theorem 1 guarantees that any causal model for these correlations requires fine-tuning.

Although the LSW inequality does not have a quantum violation for projective measurements, a very similar analysis would hold for the slightly more complex scenario of Klyachko *et al.* [30], involving five observables with binary outcomes and a cyclical compatibility structure. We leave this as an exercise for the reader.

## V. RELATION TO PREVIOUS WORK

The present work was inspired by Ref. [20]; however, there are some important differences. In Ref. [20], it was shown that any causal model for a Bell scenario that satisfies the no-signaling and no-fine-tuning assumptions, plus two extra assumptions of marginal-setting independence and local-setting dependence, is factorizable. As the present derivation does not make those extra assumptions, it generalizes the result of Ref. [20] already for Bell scenarios.

Local-setting dependence is the assumption that the local outcome is *not* independent of the local setting, i.e., that it is not the case that  $(A \perp\!\!\!\perp X)$  or  $(B \perp\!\!\!\perp Y)$ . This assumption rules out standard examples of Bell-inequality violation with unbiased outcomes. Marginal-setting independence is the assumption that the settings  $X$  and  $Y$  are uncorrelated,  $(X \perp\!\!\!\perp Y)$ . While in Ref. [20] the no-signaling condition is justified by spacelike separation, here the more general no-disturbance condition is implied by measurement compatibility. In Bell scenarios, the compatibility of  $X$  and  $Y$  is guaranteed because they are chosen from disjoint sets of measurements  $\mathcal{M}_A$  and  $\mathcal{M}_B$ , but in general contextuality scenarios,  $X$  and  $Y$  are chosen from the same set  $\mathcal{M}$  and are not, in general, independent. Thus, the result of Ref. [20] cannot be applied to contextuality scenarios.

Note that the Bell inequalities implied by the present result are the usual constraints on conditional probabilities  $\mathcal{P}(AB|XY)$ . In some contexts, such as that of biased nonlocal games [21], one may be interested in constraints on the unconditional  $\mathcal{P}(ABXY)$ . The class of classical models considered in that paper is one where the usual constraints on conditional probabilities apply but where the joint probabilities for  $X$  and  $Y$  are allowed to be arbitrary. Therefore,  $\mathcal{P}(ABXY) = \mathcal{P}(XY)\mathcal{P}(AB|XY) = \mathcal{P}(XY)\sum_{\Lambda} P(\Lambda)P(A|X\Lambda)P(B|Y\Lambda)$ . It was shown in Ref. [21] that quantum correlations can violate inequalities implied by this model for some values of  $\mathcal{P}(XY)$ , even in cases where the conditional probabilities do not violate a Bell inequality. Since here we make no assumption about the joint probabilities  $\mathcal{P}(XY)$ , the set of classical models above can also be derived from the assumption of no fine-tuning, and therefore, classical causal models that reproduce the quantum violations of those constraints also

require fine-tuning—even when they do not allow for the violation of a Bell inequality.

## VI. CONCLUSION

In summary, we have derived KS inequalities as a consequence of the principle of no fine-tuning for causal models of phenomena that satisfy the no-disturbance conditions. This result unifies Bell nonlocality (where the no-disturbance condition corresponds to the no-signaling condition) and KS contextuality as violations of classical causality. Remarkably, unlike all other derivations of Bell inequalities, this result needs no assumption related to the independence of settings, such as free choice,  $\Lambda$  independence, or marginal-setting independence.

One could object that to achieve the no-disturbance conditions in general contextuality scenarios, one requires perfectly compatible measurements and that this idealization makes it inapplicable to real experimental tests. This is true, but it is no worse than the problem faced by all standard derivations of KS inequalities, as discussed in Ref. [7], where the assumption of outcome determinism is incompatible with realistic unsharp measurements. The advantage of the present work is that, at least for idealized phenomena, it allows the derivation of KS inequalities from causality principles alone, without the extra assumption of outcome determinism. Thus, it allows for the conclusion that those causal principles must be revised in light of the (idealized) predictions of quantum theory, whereas no such conclusion can be reached with the usual derivation, even in the idealized case.

Furthermore, the present derivation suggests a path for an experimentally robust generalization. Given a measure of causal connection [35,36], we can propose a generalized principle of no fine-tuning: A causal model should not allow causal connections stronger than needed to explain the observed deviations from the no-disturbance condition. It would be interesting to determine whether testable constraints can be derived this way. Another interesting question is whether the proof can be extended to arbitrary numbers of measurements per context.

From the point of view of applications, we may understand fine-tuning as a kind of “resource waste,” postulating causal links that are not directly observed in the phenomena but that are washed out by our ignorance of underlying parameters. It is plausible to conjecture that quantum causal models [14–18], on the other hand, can avoid fine-tuning in explaining quantum correlations. If we think of classical causal models as classical simulations of quantum phenomena, the present result implies that they must necessarily waste resources via fine-tuning. A quantification of this intuition could potentially provide a novel picture to explain the power of contextuality as a resource for quantum computation.

Finally, it would be interesting to determine whether the generalized notions of noncontextuality given by the formalism of Spekkens [3] can also be understood as arising from the no-fine-tuning condition.

## ACKNOWLEDGMENTS

This research was supported by Grant No. FQXi-RFP-610 1504 from the Foundational Questions Institute Fund (fqxi.org) at the Silicon Valley Community Foundation, and by the Australian Research Council Centre of Excellence Project No. CE170100012. The author would like to acknowledge helpful discussions with Robert Spekkens, Nadish de Silva, Howard Wiseman, Rafael Chaves, Leandro Aolita, Sally Shrapnel, Fabio Costa, David Schmid, and Adán Cabello, and constructive comments from two anonymous referees.

## APPENDIX: PROOF OF THEOREM 1

*Proof.*—We prove by exhaustion that all DAGs that do not require fine-tuning to explain the no-disturbance conditions lead to factorizability and, thereby, to KS noncontextuality.

First, note that the no-disturbance conditions, together with the assumption of no fine-tuning, imply that every DAG  $\mathcal{G}$  for a no-disturbance phenomenon  $\mathcal{P}$  must satisfy the  $d$ -separation conditions  $(A \perp\!\!\!\perp Y|X)_d$  and  $(B \perp\!\!\!\perp X|Y)_d$ . We thus proceed by excluding every DAG that does not satisfy these conditions and by showing that all remaining DAGs imply factorizability.

The class of DAGs we need to consider are those that include latent variables as common causes for observable variables or direct causal connections between variables. There is no point considering latent variables as intermediaries between variables, or as common effects of variables, since adding those has no effect on the allowed probability distributions over the observable variables.

To aid the proofs, I introduce the graphical notation in Figs. 4–7 to represent sets of causal connections.

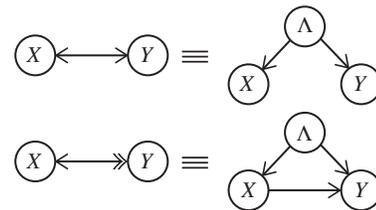


FIG. 4. Shortcut graphical notations for causal connections between  $X$  and  $Y$ .

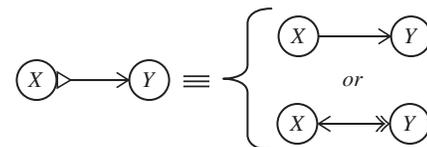


FIG. 5. Shortcut graphical notation for a direct cause from  $X$  to  $Y$  with or without a common cause.

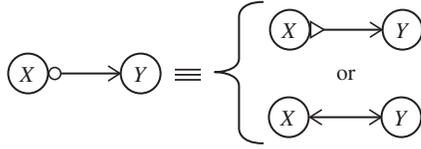


FIG. 6. Shortcut graphical notation for any causal link with no direct cause from  $Y$  to  $X$ .

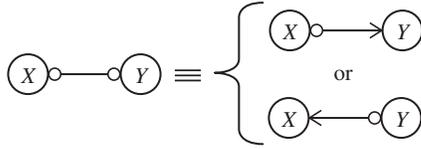


FIG. 7. Shortcut graphical notation for any causal link between  $X$  and  $Y$ .

As a reminder, the  $d$ -separation condition  $(A \perp\!\!\!\perp Y|X)_d$  means that  $X$  “blocks” all paths  $p$  from  $A$  to  $Y$ . A path is blocked by  $X$  if and only if (i) it contains a chain or a fork with  $X$  in the middle node, or (ii) it contains an inverted fork such that  $X$  is not the middle node nor a descendant of it. The intuition behind these rules is that in case (i),  $X$  is a common cause or an intermediate cause between its adjacent variables, and thus conditioning on  $X$  eliminates the correlations established by this causal path, thereby “blocking” it. In case (ii),  $X$  is a common effect of its adjacent variables, and thus conditioning on  $X$  can render them correlated. Therefore, from this  $d$ -separation condition, we eliminate all graphs that contain one or more paths between  $A$  and  $Y$  that are not blocked by  $X$ , and likewise for  $(B \perp\!\!\!\perp X|Y)_d$ .

*Step 1:* From the  $d$ -separation condition  $(A \perp\!\!\!\perp Y|X)_d$ , we can exclude any direct causal link or common cause between  $A$  and  $Y$  (i.e., all edges of the kind shown in Fig. 7). Likewise, from  $(B \perp\!\!\!\perp X|Y)_d$ , we can exclude any direct causal link or common cause between  $B$  and  $X$ . Taken together, these exclude common causes between any three or all four of the variables. We are left with the possibility of any causal links between the pairs  $\{A, B\}$ ,  $\{X, A\}$ ,  $\{Y, B\}$ , and  $\{X, Y\}$ , as shown in Fig. 8. Note that I do not assume *a priori* that common causes can only act between at most two variables—this is a consequence of no fine-tuning.

*Step 2a:* Next, we exclude a direct causal link from  $A$  to  $B$  (with or without a common cause between those two variables). First, note that the assumption of such a link excludes any causal link between  $A$  and  $X$ , as those would violate  $(B \perp\!\!\!\perp X|Y)_d$ , and a direct link from  $B$  to  $Y$ , as this would violate  $(A \perp\!\!\!\perp Y|X)_d$ . The remaining class of graphs compatible with a direct link from  $A$  to  $B$  can now have any connection between  $X$  and  $Y$  plus any link with no direct cause from  $B$  to  $Y$  (Fig. 9).

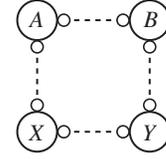


FIG. 8. Remaining class of DAGs after step 1. Dashed lines represent a connection of the type indicated or its absence.

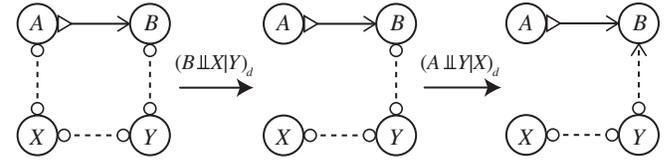


FIG. 9. DAG elimination from  $d$ -separation in Step 2a. The diagrams in this and the following figures should be read as follows. Each graph represents the set of all DAGs that contain only causal connections compatible with each shortcut notation. Solid lines require a compatible causal connection, and dashed lines are compatible with no connection. On each arrow between graphs, we eliminate all graphs that are incompatible with the  $d$ -separation condition indicated.

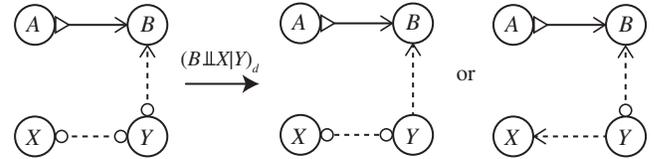


FIG. 10. DAG elimination from  $d$ -separation in step 2b.

*Step 2b:* We now exclude a common cause between  $B$  and  $Y$ , acting together with a direct link from  $X$  to  $Y$  and/or a common cause between  $X$  and  $Y$ ; those graphs would violate  $(B \perp\!\!\!\perp X|Y)_d$  as they are colliders. There are now two classes of graphs compatible with a direct link between  $A$  and  $B$ : (i) any link between  $X$  and  $Y$  and a direct link from  $Y$  to  $B$ , or (ii) a direct link from  $Y$  to  $X$  plus any link with no direct cause from  $B$  to  $Y$  (Fig. 10).

*Step 2c:* We proceed to show that all phenomena compatible with the two classes of graphs remaining after Step 2b are factorizable. To see this, first note that both classes of graphs (i) and (ii) above respect the following  $d$ -separation conditions:  $(AB \perp\!\!\!\perp X|Y)_d$  and  $(A \perp\!\!\!\perp Y)_d$ . This means that all distributions compatible with those graphs must respect  $P(AB|XY) = P(AB|Y)$  and  $P(A|Y) = P(A)$ . The first conditional independence implies that the joint distribution of  $A$  and  $B$  does not depend on the choice of measurement  $X$ , which intuitively should imply that this phenomenon cannot be contextual. To see this formally, note that from the definition of conditional probability and the two equations above, we get  $P(AB|XY) = P(B|AY)P(A|Y) = P(B|AY)P(A)$ . Now, let  $\Lambda$  be a

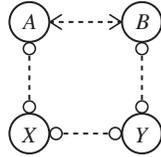


FIG. 11. Remaining class of DAGs after step 2.

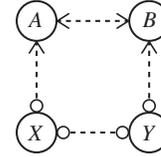


FIG. 13. Remaining class of DAGs after step 3.

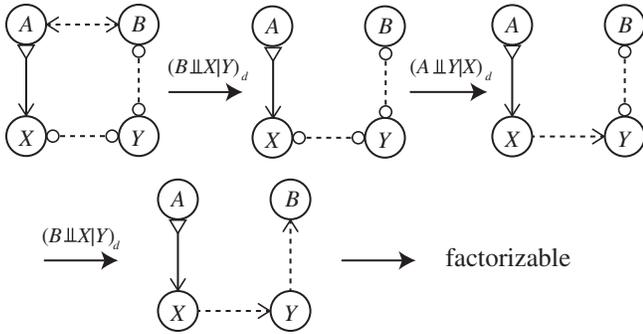


FIG. 12. Elimination of DAGs in step 3.

variable that determines  $A$ , so  $P(A) = \sum_{\Lambda} P(\Lambda)P(A|\Lambda)$  and  $P(B|AY) = P(B|Y\Lambda)$ . Then,  $P(AB|XY) = \sum_{\Lambda} P(\Lambda) \times P(A|\Lambda)P(B|Y\Lambda)$ , which is a factorizable model with no dependence on  $X$ .

This concludes the part of the proof excluding a direct causal link from  $A$  to  $B$ . By symmetry, we exclude any direct causal link from  $B$  to  $A$ . The remaining class of graphs now can have a common cause between  $A$  and  $B$  and any causal link between the pairs  $\{X, A\}$ ,  $\{Y, B\}$ , and  $\{X, Y\}$  (Fig. 11).

*Step 3:* We now proceed to exclude, from the remaining graphs, any direct cause from  $A$  to  $X$  (a retrocausal model). First, we see that, assuming such a link, a common cause between  $A$  and  $B$  is excluded from  $(B \perp X|Y)_d$ . Next, any link between  $X$  and  $Y$  except  $X \rightarrow Y$  is excluded from  $(A \perp Y|X)_d$ . Finally, any link between  $Y$  and  $B$  except  $Y \rightarrow B$  is excluded by  $(B \perp X|Y)_d$ .

The remaining class of graphs, as shown in Fig. 12, has four possible links: a common cause between  $A$  and  $X$ ,  $A \rightarrow X$ ,  $X \rightarrow Y$ , and  $Y \rightarrow B$ . This implies that  $(B \perp AX|Y)$  and  $(A \perp Y|X)$ . Thus,  $P(AB|XY) = P(B|AXY)P(A|XY) = P(A|X)P(B|Y)$ , which is trivially factorizable. By symmetry, we eliminate any direct cause from  $B$  to  $Y$ .

*Step 4:* After Step 3, we are left with the following class of graphs (Fig. 13): a possible common cause (let us call it  $\Lambda$ ) between  $A$  and  $B$ , any link between  $X$  and  $Y$ , no direct cause  $A \rightarrow X$ , and no direct cause  $B \rightarrow Y$ .

In the next step, illustrated in Fig. 14, we exclude, from  $(A \perp Y|X)_d$ , any common cause  $A \leftrightarrow X$  acting together with  $X \leftrightarrow Y$  and/or  $X \leftarrow Y$ . Likewise, from  $(B \perp X|Y)_d$ , we exclude any common cause  $B \leftrightarrow Y$  acting together with  $X \leftrightarrow Y$  and/or  $X \rightarrow Y$ .

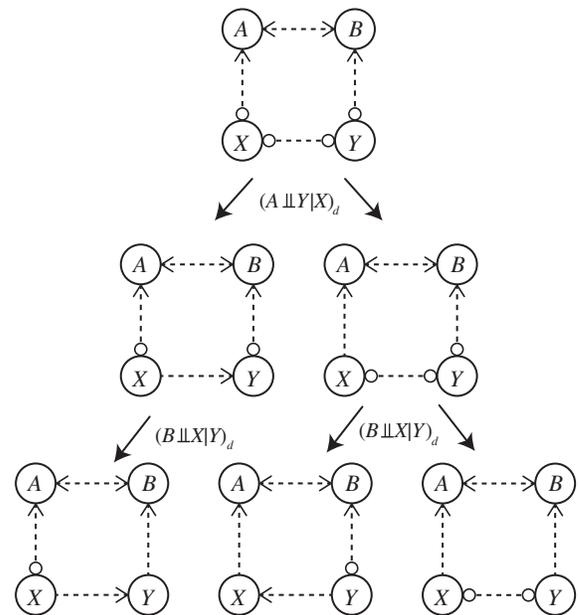


FIG. 14. DAG elimination in step 4. In step 5 we show that the three remaining classes of DAGs are factorizable.

*Step 5:* We are finally left with three remaining classes of graphs. All three classes allow for  $A \leftrightarrow B$ ,  $X \rightarrow A$ , and  $Y \rightarrow B$ , and, respectively, (i)  $X \leftrightarrow A$ ,  $X \rightarrow Y$ , (ii)  $Y \leftrightarrow B$ ,  $X \leftarrow Y$ , and (iii) any link between  $X$  and  $Y$ . All of these have  $\Lambda$  as a free variable and imply the conditional independences  $(A \perp BY|X\Lambda)$  and  $(B \perp AX|Y\Lambda)$ . Thus,  $P(AB|XY) = \sum_{\Lambda} P(\Lambda)P(A|BXY\Lambda)P(B|XY\Lambda) = \sum_{\Lambda} P(\Lambda)P(A|X\Lambda)P(B|Y\Lambda)$ , which is of factorizable form. This completes the proof.

[1] S. Kochen and E.P. Specker, *The Problem of Hidden Variables in Quantum Mechanics*, J. Math. Mech. **17**, 59 (1967).  
 [2] M. Howard, J. Wallman, V. Veitch, and J. Emerson, *Contextuality Supplies the “Magic” for Quantum Computation*, Nature (London) **510**, 351 (2014).  
 [3] R.W. Spekkens, *Contextuality for Preparations, Transformations, and Unsharp Measurements*, Phys. Rev. A **71**, 052108 (2005).  
 [4] S. Abramsky and A. Brandenburger, *The Sheaf-Theoretic Structure of Non-locality and Contextuality*, New J. Phys. **13**, 113036 (2011).

- [5] A. Cabello, S. Severini, and A. Winter, *Graph-Theoretic Approach to Quantum Correlations*, *Phys. Rev. Lett.* **112**, 040401 (2014).
- [6] A. Acín, T. Fritz, A. Leverrier, and A. B. Sainz, *A Combinatorial Approach to Nonlocality and Contextuality*, *Commun. Math. Phys.* **334**, 533 (2015).
- [7] R. W. Spekkens, *The Status of Determinism in Proofs of the Impossibility of a Noncontextual Model of Quantum Theory*, *Found. Phys.* **44**, 1125 (2014).
- [8] J. Barrett and A. Kent, *Non-contextuality, Finite Precision Measurement and the Kochen-Specker Theorem*, *Stud. Hist. Phil. Mod. Phys.* **35**, 151 (2004).
- [9] J. S. Bell, *On the Einstein-Podolsky-Rosen Paradox*, *Physics* **1**, 195 (1964).
- [10] H. M. Wiseman and E. G. Cavalcanti, *Causarum Investigatio and the Two Bell's Theorems of John Bell*, in *Quantum [Un]Speakables II—Half a Century of Bell's Theorem*, edited by R. Bertlmann and A. Zeilinger (Springer, New York, 2015), Chap. 6, pp. 119–142.
- [11] J. Pearl, *Causality: Models, Reasoning and Inference* (Cambridge University Press, Cambridge, England, 2000).
- [12] H. Reichenbach, *The Direction of Time* (University of Los Angeles Press, Berkeley, 1956).
- [13] E. G. Cavalcanti and R. Lal, *On Modifications of Reichenbach's Principle of Common Cause in Light of Bell's Theorem*, *J. Phys. A* **47**, 424018 (2014).
- [14] J. Henson, R. Lal, and M. F. Pusey, *Theory-Independent Limits on Correlations from Generalized Bayesian Networks*, *New J. Phys.* **16**, 113043 (2014).
- [15] J. Pienaar and Č. Brukner, *A Graph-Separation Theorem for Quantum Causal Models*, *New J. Phys.* **17**, 073020 (2015).
- [16] R. Chaves, C. Majenz, and D. Gross, *Information-Theoretic Implications of Quantum Causal Structures*, *Nat. Commun.* **6**, 5766 (2015).
- [17] F. Costa and S. Shrapnel, *Quantum Causal Modelling*, *New J. Phys.* **18**, 063032 (2016).
- [18] J.-M. A. Allen, J. Barrett, D. C. Horsman, C. M. Lee, and R. W. Spekkens, *Quantum Common Causes and Quantum Causal Models*, *Phys. Rev. X* **7**, 031021 (2017).
- [19] E. G. Cavalcanti, *Bell's Theorem and the Measurement Problem: Reducing Two Mysteries to One?*, *J. Phys. Conf. Ser.* **701**, 012002 (2016).
- [20] C. J. Wood and R. W. Spekkens, *The Lesson of Causal Discovery Algorithms for Quantum Correlations: Causal Explanations of Bell-Inequality Violations Require Fine-Tuning*, *New J. Phys.* **17**, 033002 (2015).
- [21] T. Lawson, N. Linden, and S. Popescu, *Biased Nonlocal Quantum Games*, [arXiv:1011.6245](https://arxiv.org/abs/1011.6245).
- [22] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, *Bell Nonlocality*, *Rev. Mod. Phys.* **86**, 419 (2014).
- [23] Note that pairwise compatibility in a set of measurements  $\mathcal{M}$  does not, in general, imply that all measurements in  $\mathcal{M}$  are jointly measurable [24].
- [24] 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).
- [25] A. Fine, *Hidden Variables, Joint Probability, and the Bell Inequalities*, *Phys. Rev. Lett.* **48**, 291 (1982).
- [26] S. Abramsky and L. Hardy, *Logical Bell Inequalities*, *Phys. Rev. A* **85**, 062114 (2012).
- [27] I. Pitowsky, *Quantum Probability—Quantum Logic*, Lecture Notes in Physics (Springer-Verlag, Berlin, 1989).
- [28] A note on terminology: Some authors (e.g., Ref. [29]) use “NC inequalities” [30,31] to refer to what I call KS inequalities, and they use the latter term to refer to inequalities that, apart from factorizability, make the extra assumption that outcomes assigned by the hidden variables satisfy the same constraints as the corresponding quantum operators (e.g., that outcomes corresponding to orthogonal projectors are mutually exclusive) [32,33]. However, for every KS inequality (in the sense of those authors) that has a quantum violation, there exists a NC inequality that is violated by the same statistics. The situation is analogous to that of the original Bell inequality [9], which apart from a local hidden variable model, also assumes a quantum prediction of perfect correlations. The term Bell inequality, however, now refers to those derived without that unnecessary assumption. Likewise, I advocate the use of KS inequalities to refer to those derived without assuming the quantum predictions.
- [29] I. Bengtsson, K. Blanchfield, and A. Cabello, *A Kochen-Specker Inequality from a SIC*, *Phys. Lett. A* **376**, 374 (2012).
- [30] A. A. Klyachko, M. A. Can, S. Binicioglu, and A. S. Shumovsky, *Simple Test for Hidden Variables in Spin-1 Systems*, *Phys. Rev. Lett.* **101**, 020403 (2008).
- [31] A. Cabello, *Experimentally Testable State-Independent Quantum Contextuality*, *Phys. Rev. Lett.* **101**, 210401 (2008).
- [32] C. Simon, Č. Brukner, and A. Zeilinger, *Hidden-Variable Theorems for Real Experiments*, *Phys. Rev. Lett.* **86**, 4427 (2001).
- [33] J.-A. Larsson, *A Kochen-Specker Inequality*, *Europhys. Lett.* **58**, 799 (2002).
- [34] D. Dürr, S. Goldstein, and N. Zanghí, *Quantum Equilibrium and the Origin of Absolute Uncertainty*, *J. Stat. Phys.* **67**, 843 (1992).
- [35] D. Janzing, D. Balduzzi, M. Grosse-Wentrup, and B. Schölkopf, *Quantifying Causal Influences*, *Ann. Stat.* **41**, 2324 (2013).
- [36] R. Chaves, R. Kueng, J. B. Brask, and D. Gross, *Unifying Framework for Relaxations of the Causal Assumptions in Bell's Theorem*, *Phys. Rev. Lett.* **114**, 140403 (2015).