

Witnessing Nonclassicality in a Causal Structure with Three Observable Variables


Pedro Lauand^{1,*}, Davide Poderini,² Ranieri Nery,² George Moreno³, Lucas Pollyceno¹,
Rafael Rabelo¹ and Rafael Chaves^{2,4}

¹*Instituto de Física “Gleb Wataghin”, Universidade Estadual de Campinas, Campinas 130830-859, Brazil*

²*International Institute of Physics, Federal University of Rio Grande do Norte, Natal 59078-970, Brazil*

³*Departamento de Computação, Universidade Federal Rural de Pernambuco, Recife, Pernambuco 52171-900, Brazil*

⁴*School of Science and Technology, Federal University of Rio Grande do Norte, Natal, Brazil*

 (Received 8 December 2022; revised 18 January 2023; accepted 20 March 2023; published 20 April 2023)

Seen from the modern lens of causal inference, Bell’s theorem is nothing other than the proof that a specific classical causal model cannot explain quantum correlations. It is thus natural to move beyond Bell’s paradigmatic scenario and consider different causal structures. For the specific case of three observable variables, it is known that there are three nontrivial causal networks. Two of those are known to give rise to quantum nonclassicality: the instrumental and the triangle scenarios. Here we analyze the third and remaining one, which we name the Evans scenario, akin to the causal structure underlying the entanglement-swapping experiment. We prove a number of results about this elusive scenario and introduce new and efficient computational tools for its analysis that can also be adapted to deal with more general causal structures. We do not solve its main open problem—whether quantum nonclassical correlations can arise from it—but give a significant step in this direction by proving that postquantum correlations, analogous to the paradigmatic Popescu-Rohrlich box, do violate the constraints imposed by a classical description of the Evans causal structure.

DOI: [10.1103/PRXQuantum.4.020311](https://doi.org/10.1103/PRXQuantum.4.020311)

I. INTRODUCTION

Bell’s theorem [1] is a cornerstone of quantum theory, having far-reaching implications for its foundations as well as in applications for quantum information processing [2]. The violation of a Bell inequality provides device-independent [3] proof of the incompatibility of classical and quantum predictions, that is, solely based on the causal assumptions of an experiment and agnostic of any internal mechanisms of the involved physical apparatuses. More generally, it evidences the need for a genuine notion of quantum causal models [4–12] in order to explain the correlations we observe in nature.

Importantly, the mismatch between classical and quantum causal predictions can be generalized to causal structures beyond that in the paradigmatic Bell scenario. Motivated by the steady progress on quantum networks [13], there have been a number of results [13–21] proving

that correlations across the distant parties of causal networks composed of independent sources can also exhibit nonclassical behavior, as already proven in a number of experiments [22–28]. In particular, quantum networks allow for a novel form of nonlocality that, as opposed to Bell’s theorem, does not require the need for measuring different observables [15,18,29]. In parallel, temporal scenarios based on causal structures involving communication between the parties [30,31] have also provided a fruitful path for understanding the role of causality in quantum theory, for instance, showing that interventions—a central concept in the field of causal inference—are able to reveal nonclassicality in situations where Bell-like tests, based on observations, would simply fail [32,33].

It is thus natural to ask what causal structures can lead to nonclassical behavior, a question that has remained elusive for two main reasons. The first is the fact that the number of possible causal structures increases very rapidly. Even to prove their equivalence classes—that is, which causal structures can give rise to the same set of observed correlations—has only been solved up to three observable variables [34,35]. The other hurdle stems from the nonconvex nature of the set of correlations permitted by general causal models [36,37]. In spite of the number of complementary approaches developed in recent years [38–44],

*p223457@dac.unicamp.br

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

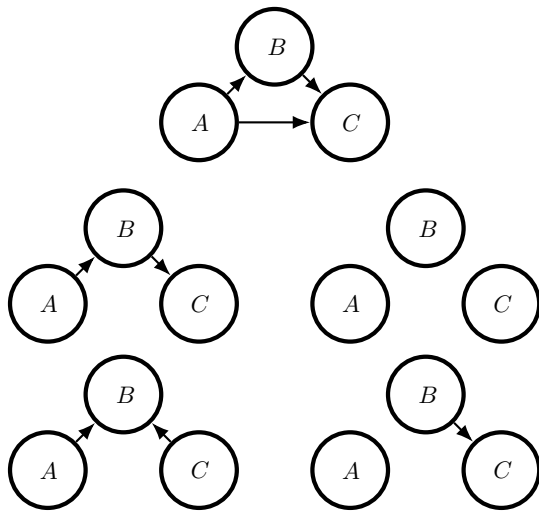


FIG. 1. Possible tripartite causal structures with no latent variables. Of the possible eight classes of causal structures with three observable nodes, five are represented by structures with no latent variables.

their practical use is still limited to a few cases of interest, which furthermore have to be evaluated on a case-to-case basis.

In the case of three observable nodes, it has been proved that there are a total of eight inequivalent classes of causal structures; see Figs. 1 and 2. From those, only three involve latent variables—that in a quantum description could be represented by entangled states—and thus lead to correlations without a classical analog. Of these three, Fig. 2(a) corresponds to the instrumental scenario [31,45] and Fig. 2(b) to the triangle scenario [15,18], bounded by Bell-like inequalities that can be violated with the help of entanglement, proving their nonclassical nature. For the third causal structure, depicted in Fig. 2(c) and to which we refer to as the Evans causal structure [34], it is not known whether it can lead to nonclassical correlations. We show that nonclassical correlations reminiscent of Popescu-Rohrlich (PR) boxes [46] can violate the constraints implied by a classical description of this causal structure. The quantum violation of such bounds remains an open problem; however, as we show, a natural class of quantum correlations does have a classical explanation in such a scenario.

The paper is organized as follows. In Sec. II we introduce causal structures, their representation as directed graphs as well as a brief discussion of their equivalence classes. In Sec. III we prove some general results for the quantum and classical compatible distribution in the Evans scenario, using its similarity with the bilocality scenario. We further prove that, despite their similarities, a natural set of correlations in the bilocality scenario, involving measurements in the maximally entangled basis, does have a classical explanation in the Evans scenario, pointing out

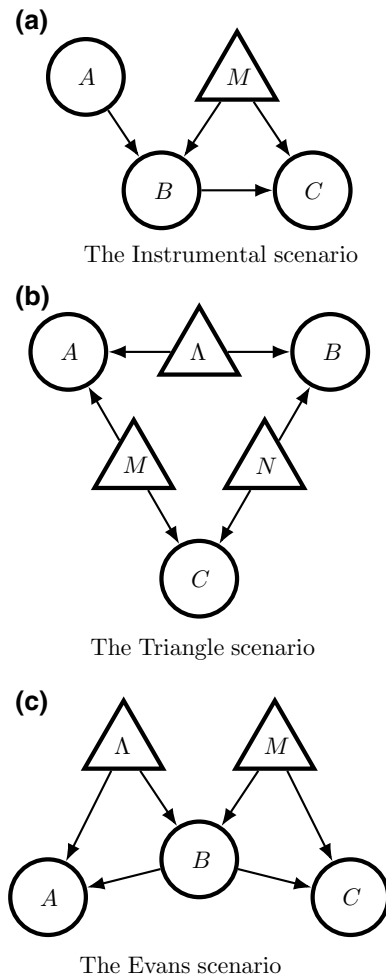


FIG. 2. Possible tripartite causal structures with latent variables. There are a total of eight inequivalent classes of causal structures that can involve three observable nodes. Of these, only three of them contain at least one latent variable, a necessary condition to display a classical-quantum gap, that is, quantum correlations that cannot be explained by a classical causal model. Two of them are known to have quantum-classical gap: the Instrumental (a) and the triangle (b). Quantum violations for Evans’s scenario (c) remain an open problem.

that the possible existence of genuine quantum distributions requires a more subtle approach. In Secs. IV and V we discuss two approaches to the causal compatibility problem: (i) nonconvex quadratic optimization and (ii) the inflation technique [47] augmented with e-separation [48], using both of them to demonstrate the nonclassicality of a postquantum distribution (similar to a PR box) in the Evans scenario. Finally, in Sec. VI we discuss methods to derive conditions valid regardless of the nature of the latent sources (classical, quantum, or postquantum), which can effectively test the topology of the Evans causal structure.

II. DAGS AND THEIR INEQUIVALENCE CLASSES

A causal structure is represented by a directed acyclic graph (DAG) \mathcal{G} that consists of a finite set of nodes $N_{\mathcal{G}}$ and a set of directed edges $E_{\mathcal{G}} \subseteq N_{\mathcal{G}} \times N_{\mathcal{G}}$. These graphs need some distinction among the vertices to clarify if a node in the graph represents either an observable or an unobserved (or latent) variable. Graphically, we use the circles to represent observable variables and triangles for latent variables (see, for example, Fig. 2), and we denote the latter with greek letters. We can define the concept of causal parents $\text{Pa}(A)$ of a given variable A in \mathcal{G} as the set of nodes sharing incoming edges with A , i.e., $\text{Pa}(A) := \{X \in N_{\mathcal{G}} | X \rightarrow A\}$. Similarly, the children of a node A are defined as those nodes at which edges originating at A terminate, i.e., $\text{Ch}(A) := \{X \in N_{\mathcal{G}} | A \rightarrow X\}$.

Every causal structure specifies a causal model for its nodes via a family of causal parameters. The causal parameters specify, for each node $X \in N_{\mathcal{G}}$, the conditional probability distribution $p_{X|\text{Pa}(X)}(x|y)$ over the values of the random variable X , given the values of the variables in $y \in \text{Pa}(X)$. If we have exogenous variables, i.e., $\text{Pa}(X) = \emptyset$, the conditional distribution becomes an unconditioned distribution over variable X .

Now, consider a DAG \mathcal{G} with $L_{\mathcal{G}} = \{\Lambda_1, \dots, \Lambda_I\}$, the set of latent variables, and $O_{\mathcal{G}} = \{A_1, \dots, A_J\}$, all the observable nodes. A causal model specifies a joint distribution of all variables in the causal structure via

$$p(\lambda_1, \dots, \lambda_I, a_1, \dots, a_J) = \prod_{v \in N_{\mathcal{G}}} p(v | \text{Pa}(v)). \quad (1)$$

We can partition $N_{\mathcal{G}} = O_{\mathcal{G}} \cup L_{\mathcal{G}}$ and rewrite Eq. (1) as a product of latent and observable causal parameters. Giving an observable joint distribution $p(a_1, \dots, a_J)$ over $A_i \in O_{\mathcal{G}} \subset N_{\mathcal{G}}$ satisfying the *global Markov condition*, that is, it admits the following decomposition:

$$p(a_1, \dots, a_J) = \sum_{\lambda_1, \dots, \lambda_I} \prod_{\lambda_i \in L_{\mathcal{G}}} p(\lambda_i | \text{Pa}(\lambda_i)) \times \prod_{a_j \in O_{\mathcal{G}}} p(a_j | \text{Pa}(a_j)). \quad (2)$$

Note that we are interested in the case of exogenous latent variables, i.e., $\text{Pa}(\lambda_i) = \emptyset$.

Another important concept is the *d-separation* criterion. The d-separation is a graphical method for deciding, from a given directed graph \mathcal{G} , whether a set A of variables is independent of another set B , given a third set C . The idea is to associate “dependence” with “connectedness,” i.e., the existence of a connecting path, and “independence” with “unconnectedness” or “separation.”

We first define the concept for an *undirected path* in a DAG and a *collider*. An undirected path $P \subset N_{\mathcal{G}}$ is a set of consecutive nodes of \mathcal{G} connected by edges, regardless

of their direction. A collider is a variable that has more than one causal parent but no children in P ; equivalently, a collider can be thought of as an inverted *fork*. A fork is a node that only has outgoing edges in P . When a variable is neither a fork nor a collider, we have a *chain*, i.e., the variable has incoming and outgoing edges in P .

Definition 1: Let \mathcal{G} be a given causal structure, and let A , B , and C be disjoint sets of variables. Consider P_{AB} the path that connects the variables from A to the variables from B . We say that A and B are d-separated by C if any of the following conditions hold:

- (a) P_{AB} contains a collider $m (\rightarrow m \leftarrow)$ such that $m \notin C$, and all $y \in \text{Ch}(m)$ also satisfy $y \notin C$;
- (b) P_{AB} contains a fork $m (\leftarrow m \rightarrow)$ such that $m \in C$;
- (c) P_{AB} contains a chain variable $m (\rightarrow m \rightarrow$ or $\leftarrow m \leftarrow)$ such that $m \in C$.

We denote this by $A \perp_d B | C$.

Note that the conditions in the definition above are graphical, i.e., they depend only on the structure of \mathcal{G} , and allow us to derive nontrivial conditions on the correlations. It is possible to show that d-separation is equivalent to the nodes satisfying the conditional independence relation $A \perp B | C$, that is, $p(ab|c) = p(a|c)p(b|c)$ [49]. Importantly, any probability distribution over $O_{\mathcal{G}}$ is compatible with the global Markov model only if it respects all d-separation conditions of \mathcal{G} .

Crucially, the compatibility definitions described above are valid only if we assume that the latent variables behave like classical systems. If we are free to identify the unobserved nodes as sources of quantum correlations, the set of quantum compatible distributions is, for certain scenarios, strictly larger than the classical one. Specifically, we can define a quantum compatible distribution for a DAG \mathcal{G} as that described by the following strategy.

1. To each latent variable $\Lambda \in L_{\mathcal{G}}$, we associate a quantum state described by the density operator $\rho_{\Lambda} \in \mathcal{L}(\mathcal{H}_{\Lambda})$.
2. To each observed variable $A \in O_{\mathcal{G}}$, we associate a positive operator-valued measure (POVM) measurement described by operators, $\{E_{\text{Pa}^O(a)}^a\}_a$, dependent on the outcome of its observed parents $\text{Pa}^O(a)$, and acting nontrivially only on the space $\mathcal{H}_{\text{Pa}^L(A)}$ of all the latent parents of A .
3. The distribution is obtained by the Born rule applied to the state of all latent nodes:

$$p(O_{\mathcal{G}}) = \text{tr} \left(\bigotimes_{a \in O_{\mathcal{G}}} E_{\text{Pa}^O(a)}^a \bigotimes_{\Lambda \in L_{\mathcal{G}}} \rho_{\Lambda} \right). \quad (3)$$

We denote the set of all quantum compatible distributions for a DAG \mathcal{G} as $\mathcal{Q}_{\mathcal{G}}$. In this paper, we also consider resources that may not be compatible with a quantum

description and are modeled by generalized probabilistic theories. A strategy similar to that above can be used to define the set of correlations compatible with a causal structure \mathcal{G} under scrutiny. Where we take $|\omega_\Lambda\rangle \in \Omega_\Lambda$, a generalized probabilistic theory (GPT) generalization of the quantum state ρ_Λ , and $(e_{\text{Pa}O(a)}^a | \circ | \Omega)$, GPT generalizations of the quantum measurement operator $E_{\text{Pa}O(a)}^a$ and the joint distribution over the observable variables is given by

$$p(O_{\mathcal{G}}) = \left(\prod_{a \in O_{\mathcal{G}}} (e_{\text{Pa}O(a)}^a | \circ | \Omega) \right), \quad (4)$$

where $|\Omega\rangle$ lives in some composite state space $\prod_{\Lambda \in L_{\mathcal{G}}} \Omega_\Lambda$ that contains the tensor product of the state spaces as a subspace [50]. We denote the set of all distributions compatible with a GPT as $\Omega_{\mathcal{G}}$.

Being interested in the *causal compatibility problem* (CCP), it is convenient to restrict our attention to equivalence classes of DAGs that admit different sets of classically compatible distributions, that is, we consider $\mathcal{G} \sim \mathcal{G}'$ belonging to the same class whenever $\mathcal{C}_{\mathcal{G}} = \mathcal{C}_{\mathcal{G}'}$. Note that, while DAGs belonging to the same class need to have the same number of observable nodes, latent nodes are in no way restricted. It can be shown [51] that in the case of three observable nodes there are only eight different classes of DAGs, whose representatives are shown in Figs. 1 and 2. In particular, in Fig. 2 we represent the three classes of DAGs containing latent variables, which are the only ones where a strict inclusion $\mathcal{C}_{\mathcal{G}} \subset \mathcal{Q}_{\mathcal{G}}$ is possible. As previously mentioned, Evans’s scenario, depicted in Fig. 2(c), is the only one for which this question remains an open problem.

III. EVANS’S SCENARIO

In this section, we prove a number of results about the Evans scenario. First, we show a general mapping between the Evans scenario and the much-studied bilocality scenario [14]. Second, we prove that a natural class of correlations obtained with maximally entangled states and measurements do have a classical explanation. Before that we better motivate the interest in the causal structure underlying the Evans scenario.

The Evans scenario is the simplest causal structure combining the two features that have attracted the attention of the whole community lately: communication between outcomes (as in the instrumental scenario) and independent sources (as in the triangle scenario). Moreover, the Evans’s scenario is the underlying causal structure of a paradigmatic protocol in quantum information science: the entanglement swapping [52], where a central node (B) makes a joint measurement on his share of entangled states and by communicating his measurement outcomes is able to establish entanglement between two distant parties (A and C) that have never interacted before.

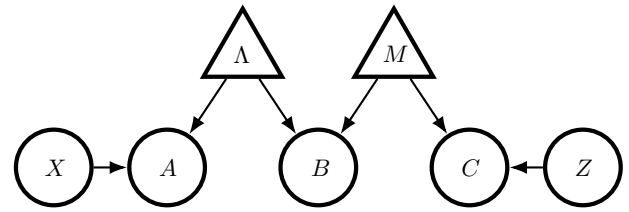


FIG. 3. Causal structure of the bilocality scenario. Two independent sources Λ and M connect nodes A and C , which are influenced by X and Z , respectively, to B .

It is also worth noting that it is not known whether the three different equivalence classes that hold classically (instrumental, triangle, and Evans) remain the same in the quantum case. As proven in Ref. [11], classically equivalent DAGs might become inequivalent if quantum states act as the common source. Nonetheless, note that any classical Bell inequality we derive for the Evans scenario will also be a valid constraint for any other member of that equivalence class. That is, even if one of these DAGs is inequivalent when considering quantum correlations, all the inequalities and methods we discuss in the rest of the paper remain valid. This means that one could generate quantum correlations using other DAGs but still use the same constraints we have used.

A. The relation between the bilocality and the Evans scenarios

In the following, we show an interesting relation between the Evans scenario to the so-called bilocal scenario [53]. This scenario, whose DAG \mathcal{G}_B is displayed in Fig. 3, presents two independent sources Λ and M shared between a central node B and two peripheral nodes A and C , and can be described by a conditional probability distribution $p(a, b, c|x, z)$. It is known [53] that, for this scenario, the classically compatible set of distributions does not coincide with the quantum set, i.e., $\mathcal{C}_B \subset \mathcal{Q}_B$.

A distribution compatible with the Evans scenario can be seen as a projection of a distribution in the bilocal scenario into the subspace $p(a, b, c|x = z = b)$. More precisely, for every distribution admitting a classical model $p_E(a, b, c)$ in the Evans scenario, there exists a realization $p_B(a, b, c|x, z)$ in the bilocal scenario such that $p_E(a, b, c) = p_B(a, b, c|x = z = b)$ and $p_B(a, b, c|x, z)$ also admits a classical model. Conversely, for every bilocal distribution $p_B(a, b, c|x, z)$, its projection, $p_E(a, b, c) := p_B(a, b, c|x = z = b)$, admits a classical model in the Evans scenario.

This simple fact allows us to derive some nontrivial conclusions about the topology of the set of classical correlations in Evans’s scenario, \mathcal{C}_e , that follow from the topology of the bilocal set. We show, in analogy to Ref. [54], that the set \mathcal{C}_e is connected and has weak star convexity for a certain subspace. A similar relation is respected under the

same projection for the case where quantum states are distributed in the network. The combination of these results shows that quantum nonbilocality is a necessary condition for a given distribution to display a classical-quantum gap in the Evans scenario.

To formalize those statements, let us consider the Evans set \mathcal{C}_e and the bilocal set \mathcal{C}_B of classical correlations. We know that, for a distribution $p(abc|xz) \in \mathcal{C}_B$, the *global Markov property* implies that

$$p(abc|xz) = \sum_{\mu,\lambda} p(\mu)p(\lambda)p(a|x\lambda)p(b|\mu\lambda)p(c|z\mu). \quad (5)$$

From this, assuming that $|X| = |Z| = |B|$, we obtain the following result.

Lemma 1. *If $p_B(abc|xz) \in \mathcal{C}_B$ then $p_E(abc) := p(abc|x = z = b) \in \mathcal{C}_e$. Conversely, if $p_E(abc) \in \mathcal{C}_e$ then there exists a $p_B(abc|xz) \in \mathcal{C}_B$ such that $p_B(abc|x = z = b) = p_E(abc)$.*

Proof. The first implication of the lemma follows directly from the definitions. Suppose now that $p_E \in \mathcal{C}_e$; then there exists λ and μ such that

$$p(abc) = \sum_{\mu,\lambda} p(\mu)p(\lambda)p(a|b\lambda)p(b|\mu\lambda)p(c|b\mu). \quad (6)$$

Now we can define

$$\tilde{p}(a|\lambda, x) := p(a|\lambda, b = x), \quad (7)$$

and similarly for $\tilde{p}(c|\mu, z)$. We can also define

$$\begin{aligned} \tilde{p}(a, b, c|x, z) &= \sum_{\mu,\lambda} p(\mu)p(\lambda) \\ & p(a|b' = x, \lambda)p(b|\mu\lambda)p(c|b' = z, \mu), \end{aligned} \quad (8)$$

where \tilde{p} is bilocal by construction and $\tilde{p}(a, b, c|x = z = b) = p(abc)$, which concludes the proof. ■

Using this mapping, we can extend some results valid for the set of bilocal distributions, to the Evans one. In particular, we can prove that \mathcal{C}_e is connected and star convex for certain subspaces, i.e., there exists a preferential point $p^* \in \mathcal{C}_e^{p_A} \subset \mathcal{C}_e$ such that, for any $p \in \mathcal{C}_e^{p_A}$, the line from p^* to p is contained in $\mathcal{C}_e^{p_A}$. Note that star convexity is a weaker notion of convexity.

Lemma 2. *It holds that \mathcal{C}_e is connected*

Proof. First, it was shown in Ref. [54] that \mathcal{C}_B is connected. More precisely, for every bilocal correlation p_B , there is

a correlation $p_\xi = \xi p_B + (1 - \xi)p_0$ that follows a continuous path connecting p_B to p_0 , where $p_0(a, b, c|x, z) = 1/|A||B||C|$ is the uniform distribution. Furthermore, p_ξ is bilocal for every ξ as it can be obtained by performing local operations on the correlation p_B .

Now, consider $p_E \in \mathcal{C}_e$. Then, by Lemma 1, there exists p_B bilocal that recovers p_E via projection; we can thus define

$$\begin{aligned} p_\xi^E(a, b, c) &= p_\xi(a, b, c|x = z = b) \\ &= \xi p_E + (1 - \xi) \frac{1}{|A||B||C|}, \end{aligned}$$

since p_ξ is bilocal; then $p_\xi^E \in \mathcal{C}_e$. This shows that every classical p_E is continuously connected to the uniform distribution, implying that \mathcal{C}_e is connected. ■

Moreover, star convexity is preserved when we project on the subspace determined by fixing a specific marginal distribution for a $p_A(a)$.

Lemma 3. *It holds that \mathcal{C}_e is star convex on the subspace of fixed p_A denoted by $\mathcal{C}_e^{p_A}$.*

Proof. We need to show that there exists a point $p^* \in \mathcal{C}_e^{p_A}$ such that any line segment between $p \in \mathcal{C}_e^{p_A}$ and p^* is inside $\mathcal{C}_e^{p_A}$.

Consider $p \in \mathcal{C}_e^{p_A}$, i.e., $p(a) = \sum_{bc} p(a, b, c) = p_A(a)$ is fixed and p has a classical decomposition in terms of λ and μ . Define $p^*(a, b, c) = p_A(a)p^*(b)p^*(c) \in \mathcal{C}_e^{p_A}$.

By providing an extra random bit ℓ distributed by the source with probability $p(\ell = 1) = 1 - p(\ell = 0) = \xi \in [0, 1]$ to Bob and Charlie, we can define the correlation p_ξ as follows. If $\ell = 1$, Bob and Charlie output their original response functions defined by p . If $\ell = 0$, they respond according to p^* . This yields $p_\xi = \xi p + (1 - \xi)p^*$. This operation can be done locally from Bob's and Charlie's labs and thus we can construct a local model for p_ξ , and $\sum_{bc} p_\xi(a, b, c) = p_\xi(a) = p_A(a)$. This shows that $p_\xi \in \mathcal{C}_e^{p_A}$. ■

The proof of Lemma 3 uses the exact same argument as in Appendix A of Ref. [54]. Lemmas 2 and 3 imply similarities between sets \mathcal{C}_e and \mathcal{C}_B . Our next result shows that there is a similar relationship at the quantum level. The sets of quantum correlations in the Evans scenario, \mathcal{Q}_e , and in the bilocal scenario, \mathcal{Q}_B , are respectively given by

$$p_E(a, b, c) = \text{tr}((\psi_{AB} \otimes \psi_{B'C})E_{a|b} \otimes E_b \otimes E_{c|b}) \in \mathcal{Q}_e$$

and

$$p_B(a, b, c|x, z) = \text{tr}((\psi_{AB} \otimes \psi_{B'C})E_{a|x} \otimes E_b \otimes E_{c|z}) \in \mathcal{Q}_B.$$

The systems are independently distributed and the $\{E_{ij}\}$ are general POVMs respecting

$$\sum_i E_{ij} = 1 \quad \text{for all } j, \quad E_{ij} \geq 0 \quad \text{for all } i, j.$$

Lemma 4. *If $p_B(abc|xz) \in \mathcal{Q}_B$ then $p_E(abc) = p(abc|x = z = b) \in \mathcal{Q}_e$. Conversely, if $p_E(abc) \in \mathcal{Q}_e$ then there exists a $p_B(abc|xz) \in \mathcal{Q}_B$ such that $p_B(abc|x = z = b) = p_E(abc)$.*

Proof. This is a consequence of making the identification $E_{a|x=k} \equiv E_{a|b=k}$, and similarly for other effects.

Indeed, if $p_B(abc|xz) \in \mathcal{Q}_B$ then

$$p_E(abc) = p(abc|x = z = b) \\ = \text{tr}((\psi_{AB} \otimes \psi_{B'C})E_{a|b} \otimes E_b \otimes E_{c|b}) \in \mathcal{Q}_e.$$

Conversely, if $p_E \in \mathcal{Q}_e$, we define $\tilde{E}_{a|x} = E_{a|b=x}$, and similarly for other parties, which gives

$$\tilde{p}(a, b, c|x, z) = \text{tr}((\psi_{AB} \otimes \psi_{B'C})E_{a|b'=x} \otimes E_b \otimes E_{c|b'=z}) \\ \in \mathcal{Q}_B.$$

This completes the proof. ■

It is important to point out that, from these results, we can prove quantum nonbilocality to be a necessary condition for the Evans scenario to display a quantum gap (a violation of a classically valid inequality).

Indeed, suppose that p_E is a quantum nonclassical distribution in the Evans scenario. By Lemma 4, there exists $p_B \in \mathcal{Q}_B$ that recovers p_E via projection. The distribution p_B is not unique, i.e., the quantified problem

$$\exists p_B \in \mathcal{Q}_B \text{ such that } p_B(a, b, c|x = z = b) = p_E(a, b, c), \quad (9)$$

may have many (infinitely many) solutions. However, there cannot be any solution $p_B \in \mathcal{C}_B$, because if there exists $p_B \in \mathcal{C}_B$ then $p_E \in \mathcal{C}_e$ by Lemma 1, a contradiction.

These relations hint at the possibility that, by starting with some nonclassical distribution in the bilocal scenario, one could possibly derive the incompatibility of the projected correlation in the Evans case, i.e., it is natural to ask whether nonbilocality could also become sufficient under some specific condition or some specific distribution. Unfortunately, we could not find quantum violations of bilocality that remained nonclassical after projection. We believe that this is due to the fact that many examples of nonbilocality rely on entanglement swapping processes that are heavily dependent on maximally entangled measurements for Bob, which, as we prove in the next section, do have a classical explanation in the Evans case.

As shown by these negative results, truly new methods are required to analyze the emergence of nonclassicality in the Evans scenario.

B. A classical model for measurements on a maximally entangled basis

Given the similarities between the bilocal and the Evans scenarios that we highlighted in the previous section, a natural quantum strategy to obtain nonclassical correlations in the Evans scenario would be to start from maximally entangled bipartite states shared by A, B and B, C and use node B to perform entanglement swapping [52], so that A and C would effectively perform measurements on a shared entangled state. Unfortunately, we show here that this kind of strategy possesses a classical model in the Evans scenario.

To make things more precise, let us assume that A, B and B, C both start with maximally entangled states of the form $|\Phi_d\rangle = (1/\sqrt{d}) \sum_i^d |ii\rangle$, where B performs the standard d -dimensional entangling measurement, in the basis $B_d = \{|\Phi_d^{n,m}\rangle\}_{m,n}$:

$$|\Phi_d^{n,m}\rangle = \frac{1}{\sqrt{d}} \sum_k^d e^{i2\pi nk/d} |k, k+m\rangle \quad (10)$$

with the sum modulo d and $n, m \in \{0, \dots, d-1\}$. Then A and C can perform an arbitrary POVM measurement on their part, $\{A_b^a\}_a$ and $\{C_b^c\}_c$, respectively, depending on the outcome $b = (n, m)$ of B . The distribution is thus given by

$$p(a, b, c) = \text{tr}(A_b^a \otimes B^b \otimes C_b^c |\Phi_d\rangle \langle \Phi_d| \otimes |\Phi_d\rangle \langle \Phi_d|). \quad (11)$$

Result 1. *For any d -dimensional POVMs $\{A_b^a\}_a$ and $\{C_b^c\}_c$ and for any dimension d , the distribution generated by the quantum strategy in Eq. (11) always has a classical realization compatible with the Evans causal structure.*

Proof. We start by noting that, independently of the value of the outcome b , we can always describe the effective state shared between A and C as the maximally entangled state $|\Phi_d\rangle$. Indeed, since, depending on b , the effectively swapped state will be one of the set $\{|\Phi_d^{n,m}\rangle\}$, we can reduce it to $|\Phi_d\rangle$ by applying a local unitary on one part. Specifically, we can repeatedly apply the operators X, Z , defined by

$$X |k\rangle = |k+1\rangle, \quad Z |k\rangle = e^{i2\pi k/d} |k\rangle. \quad (12)$$

In this way, redefining

$$\tilde{C}_{(n,m)}^c = Z^{-n} X^{-m} C_{(n,m)}^c X^m Z^n, \quad (13)$$

we obtain, conditioning on the outcome b ,

$$p(a, c|b) = \text{tr}(A_b^a \otimes \tilde{C}_b^c |\Phi_d\rangle \langle \Phi_d|) = \text{tr}((A_b^a)^T \tilde{C}_b^c) / d. \quad (14)$$

The distribution $p(a, c|b)$, as written above, is also compatible with a structure where variables A and C share a single latent variable N , and are both influenced by the same setting B (independent of N). In such a structure the quantum and classical distribution sets coincide, so we know that, in particular for Eq. (14), we can have a classical realization of the form $p(a, c|b) = \sum_v p(a|b, v)p(c|b, v)p(v)$, where $v \in \{0, \dots, d-1\}^2$. Without loss of generality, we can limit ourselves to combinations of deterministic strategies:

$$p(a, c|b) = \sum_v \delta_{af(b,v)} \delta_{cg(b,v)} p(v). \quad (15)$$

We now show that these can be converted into classical realizations compatible with the Evans scenario. Indeed, for a classical distribution in the Evans case, we would have

$$p(a, c|b) = \sum_{\lambda, \mu} \delta_{af'(b,\lambda)} \delta_{cg'(b,\mu)} p(\lambda, \mu|b), \quad (16)$$

where the term $p(\lambda, \mu|b)$ needs to satisfy $\sum_b p(\lambda, \mu|b) p(b) = p(\lambda)p(\mu)$. We note that in our case $p(b)$ is a uniform distribution, and if we also choose $p(\lambda)$ and $p(\mu)$ to be uniform, we have $p(\lambda, \mu|b) = p(b|\lambda, \mu)$, so that the above condition simply becomes the normalization of $p(b|\lambda, \mu)$. Now, if $p(v)$ is the distribution for the shared latent variable N that describes Eq. (15), we define

$$p(b|\lambda, \mu) = p(v = (\lambda + b_1, \mu + b_2)), \quad (17)$$

where $b = (b_1, b_2)$ and the addition is modulo d . We can easily see that, thanks to this definition, both $p(b|\lambda, \mu)$ and $p(\lambda, \mu|b)$ satisfy the normalization condition. Now, to make Eq. (16) coincide with Eq. (15), it only remains to define the local deterministic strategy of the former as $f'(\lambda, b) = f(\lambda - b_1, b)$ and $g'(\mu, b) = g(\mu - b_2, b)$.

Finally, we remark that if the POVMs $\{A_b^a\}_a$ and $\{C_b^c\}_c$ are of cardinality $d_A > d$ or $d_C > d$, the mutual information $I(A : C)$ is still bounded by $I(A : C) \leq \log_2(d)$. Then any such distribution can still be reproduced by strategy (16), adding local noise in A and C . ■

It is important to note that our result is valid regardless of the cardinality of the variables A and C . It was shown in Ref. [55] that, in the instrumental scenario, when the central node (in our case B) has a high cardinality, the scenario becomes less restrictive such that in the asymptotic limit of continuous B the model imposes no restriction for $p(a, b, c)$. Since the Evans model contains the instrumental model as a particular case (as will be discussed

in Sec. IV B), we can conclude that this will also be the case for the Evans model. This means that, whenever $|B| \geq |A|, |C|$, the model will be quite permissive and we can, for example, simulate perfect correlation between the variables trivially.

IV. NONCONVEX QUADRATIC PROBLEMS AND CAUSAL COMPATIBILITY

In our framework, nonconvex optimization problems arise in a very natural way: to solve the problem of deciding whether some correlation p lies inside or outside the nonconvex set of classical correlations. It has been proven in Ref. [56] that the classical set of correlations for any network structure is a semialgebraic set, i.e., it can be described by finitely many polynomial equalities and/or inequalities on the joint distribution of all observable nodes of the network. Thus, in principle, we could solve the CCP by globally solving a *polynomial optimization problem* (POP). Of course, in general, this is no easy task and there are approaches to try to solve such problems asymptotically as a hierarchy of simpler outer approximations like, for example, semidefinite programming relaxations [57] or linear programming [58].

Reformulating the Evans compatibility problem as a nonconvex maximization problem, $p(a, b, c)$ is compatible with Fig. 2(c) if and only if

$$\exists q(a_0, \dots, a_{|B|}, b, c_0, \dots, c_{|B|}) \geq 0 \quad (18a)$$

$$\text{subject to} \quad \sum_{a_i \neq b, c_j \neq b} q(a_0, \dots, a_{|B|}, b, c_0, \dots, c_{|B|}) = p(a, b, c), \quad (18b)$$

$$q(a_0, \dots, a_{|B|}, c_0, \dots, c_{|B|}) = q(a_0, \dots, a_{|B|}) q(c_0, \dots, c_{|B|}), \quad (18c)$$

$$\sum_{a_0, \dots, a_{|B|}, b, c_0, \dots, c_{|B|}} q(a_0, \dots, a_{|B|}, b, c_0, \dots, c_{|B|}) = 1, \quad (18d)$$

where

$$\begin{aligned} & q(a_0, \dots, a_{|B|}, b, c_0, \dots, c_{|B|}) \\ & := \sum_{\lambda, \mu} p(\lambda) p(\mu) p(b|\lambda, \mu) p(a_0|\lambda, b=0) \dots \\ & \quad \times p(a_{|B|}|\lambda, b=|B|) p(c_0|\mu, b=0) \dots \\ & \quad \times p(c_{|B|}|\mu, b=|B|). \end{aligned} \quad (19)$$

This problem is nonconvex, with all the conditions being at most quadratic. We can also see in Eq. (18) analogous conditions to Fine's theorem [59] for this network, i.e., necessary and sufficient conditions on the joint probability distribution of all the possible outcomes for causal compatibility. This clearly constitutes a nonconvex, quadratically

constrained feasibility test that can be globally solved with presently available solvers. Note that feasibility tests are a particular case of a constrained optimization problem, that is, when the objective function is constant. For our work, we have used the Gurobi optimizer [60], which allows us to tackle nonconvex, quadratically constrained quadratic problems (QCQPs). Contrary to many nonlinear optimization solvers, which search for locally optimal solutions, here we solve this problem by looking for global optimality [61].

Furthermore, we can rewrite this as a minimization problem introducing white noise to $p(a, b, c)$:

$$\text{minimize}_{v,q} \quad 1 - v \quad (20a)$$

$$\begin{aligned} \text{subject to} \quad & \sum_{a_i \neq b, c_j \neq b} q(a_0, \dots, a_{|B|}, b, c_0, \dots, c_{|B|}) \\ & = vp(a, b, c) + \frac{1 - v}{|A||B||C|}, \end{aligned} \quad (20b)$$

$$\begin{aligned} & q(a_0, \dots, a_{|B|}, c_0, \dots, c_{|B|}) \\ & = q(a_0, \dots, a_{|B|})q(c_0, \dots, c_{|B|}), \end{aligned} \quad (20c)$$

$$\sum_{a_0, \dots, a_{|B|}, b, c_0, \dots, c_{|B|}} q(a_0, \dots, a_{|B|}, b, c_0, \dots, c_{|B|}) = 1, \quad (20d)$$

$$q \geq 0. \quad (20e)$$

Note that this choice of noise is validated by Lemma 2, where it is shown that all Evans-compatible behaviors are continuously connected to the white noise distribution and, thus, the problem must yield a solution with $v < 1$ if p lies outside the classical set of correlations. Although the derivation of Eq. (18) does not work in general for networks with a more complex latent structure, e.g., triangle network, our tools may still tackle the CCP for these more general causal structures. Truly, Rosset *et al.* [56] showed that one may always take the cardinality of the latent variables to be finite and, by consequence, we can use the d-separation criterion to formulate a POP. This shows that we can use global nonconvex optimization as a necessary and sufficient test, up to computational tolerance, for causal compatibility.

A. Extracting infeasibility certificates from quadratic problems

If a given probability distribution $p^*(a, b, c)$ fails the CCP test, i.e., is incompatible with the classical causal model under test, we would also like to extract a witness for such nonclassical behavior, that is, a real function F such that

$$F(p(a, b, c)) \geq B^* \quad (21)$$

and $F(p^*) < B^*$, where $p(a, b, c)$ are all distributions compatible with the causal structure under scrutiny.

In the Evans scenario, for example, the compatibility conditions are stated in Eq. (18). Since any compatible $p(a, b, c)$ can be expressed as a marginalization of the joint distribution q , we can write $F = F(q)$, where the terms that will appear are the respective marginals of q optimized over the factorizing conditions, normalization and non-negativity. Given an incompatible distribution p^* , we may simply choose $F(p) = \|p - p^*\|^2$, since all terms of F are quadratic on q and this can be efficiently optimized. Note that the value $F(p) = 0$ is never possible, as this would imply that $p = p^*$, and p^* is assumed to be a nonfeasible point. Therefore, the program will return a bound B^* that is tight up to computational precision. Our witness will then be written as

$$\|p - p^*\|^2 \geq B^*,$$

and can be violated by p^* .

B. Finding inequalities for the Evans scenario starting from the instrumental scenario

The instrumental scenario, whose DAG is shown in Fig. 2(a), implies the following decomposition for any compatible distribution:

$$p(a, b, c) = \sum_{\mu} p(a)p(b|a, \mu)p(c|b, \mu)p(\mu). \quad (22)$$

We note that this can be regarded as a particular case of factorization (5) given by the Evans scenario.

The correspondence emerges when we drop the arrow from B to A and choose A to be a deterministic function of λ , that is, $p(a|b, \lambda) = \delta_{a,\lambda}$. More generally, the arrow $B \rightarrow A$ can be seen as a kind of measurement dependence in the instrumental scenario [62], since via this arrow the hidden variable μ can influence variable A . Because of that, it seems reasonable that inequalities valid for the instrumental scenario can be recycled for the Evans scenario if we consider the relaxation arising from the arrow $B \rightarrow A$. In the following, we show how the bounds of the instrumental inequalities change if applied to the Evans scenario.

Indeed, using the Gurobi optimizer, we can evaluate how the bound of known Bell-like inequalities changes when applied to the Evans scenario. Specifically, we consider Pearl's inequality [45]

$$\begin{aligned} P &= p_{ABC}(1, 0, 0)p_A(1) + p_{ABC}(1, 1, 1)p_A(0) - p_A(0)p_A(1) \\ &\leq 0 \end{aligned} \quad (23)$$

and Bonet's inequality [31,55]

$$\begin{aligned} B &= p_{ABC}(0, 1, 0)p_A(1)p_A(2) - p_{ABC}(1, 1, 0)p_A(0)p_A(2) \\ &\quad - p_{ABC}(1, 1, 1)p_A(0)p_A(2) - p_{ABC}(2, 0, 1)p_A(0)p_A(1) \\ &\quad - p_{ABC}(2, 1, 0)p_A(0)p_A(1) \\ &\leq 0. \end{aligned} \quad (24)$$

Employing the Gurobi optimizer, one can see that the inequalities are changed as

$$P \leq \frac{1}{16}, \quad B \leq \frac{1}{27}. \quad (25)$$

Unfortunately, however, we were unable to find any quantum or postquantum correlations able to violate these inequalities.

Nonetheless, this relation between the Evans and instrumental scenarios also hints at the possibility that correlations violating instrumental inequalities might also lead to nonclassical behavior in the Evans scenario. As will be detailed in the following, this is precisely the case for the PR box in the instrumental scenario with $|A| = 3$, appropriately choosing the marginal distribution of $p(a)$. We also show in Appendix A how this choice cannot be arbitrary since there are choices for $p(a)$ such that the distribution still admits a local model.

Now we show that the Bonet PR box given by

$$p_{\text{PR}}(bc|a) = \begin{cases} \frac{1}{2} & \text{if } c = b + f(a, b) \pmod{2}, \\ 0 & \text{otherwise,} \end{cases} \quad (26)$$

where $f(0, b) = 0$, $f(1, b) = b$, and $f(2, b) = b + 1 \pmod{2}$, can violate causal compatibility inequalities in the corresponding Evans scenario.

C. Violations of the Evans scenario

Solving the QCQP, we have a necessary and sufficient (up to computational precision) oracle to detect nonclassicality in the Evans scenario. By discretizing our domain of parameters p_A , and employing this oracle for each choice of p_A , the lowest visibility overall choice to certify the nonclassicality of the distribution in Eq. (26) is given by $v = 0.84$. With $p_A(0) = p_A(2) = 10/21$ and $p_A(1) = 1/21$, where the distribution under test is the convex sum, $p(a, b, c) = vp_{\text{PR}}(a, b, c) + (1 - v)1/12$, that is, a mixture of white noise with the PR box wired by asymmetrically distributed inputs.

Using the construction in Sec. IV A, we can find a witness that is generally valid for the scenario and is violated by our candidate distribution. Even though we can find incompatibility for visibilities as low as $v = 0.84$, the best choice of candidate distribution to derive the inequality is for $v = 1$. This fact becomes more intuitive when we look at this approach in a geometrical manner, since we are

looking at the largest Euclidean ball centered at the nonclassical correlation p that remains completely outside of the nonconvex (classical) set of correlations. Clearly, this is a nonconvex optimization problem and, furthermore, if we get arbitrarily close to the border of the set from outside, the radius of this ball becomes arbitrarily small, and if this radius is smaller than our computational precision, we have a problem. The best we can do is to choose the region with the largest volume, aiming to detect the largest amount of nonclassical distributions, which would correspond to the largest radius of this Euclidean ball. We find the largest radius for $v = 1$, which yields the witness

$$\text{GW} = \|p - p_{\text{PR}}\|_2^2 \geq 0.00061. \quad (27)$$

To obtain this witness, we optimized the objective function $\text{GW}(p) = \|p - p_{\text{PR}}\|_2^2$, where p_{PR} is the PR-box distribution mentioned before with the appropriate choice of p_A ; the optimal violation is the case where $p = p_{\text{PR}}$ for which we get $\text{GW}(p_{\text{PR}}) = 0$, which means a violation of $\beta = -0.00061$. We can also always show with our approach that this inequality is tight, i.e., the Gurobi program can always return a classical point that achieves the optimal objective function. More generally, this can be seen as a consequence of the Weierstrass extreme value theorem that guarantees that the continuous function $F(p) : \mathcal{C}_e \rightarrow \mathbb{R}$ can take its extreme values with points inside the compact set of classical correlations \mathcal{C}_e .

V. DERIVING A WITNESS FROM THE INFLATION TECHNIQUE AUGMENTED WITH E-SEPARATION

The inflation technique [42] is an important tool that allows us to set constraints on the correlations that can arise in any network (up to computational complexity issues). Intuitively, we are concerned with the hypothetical situation where one has access to multiple copies of the sources and measurement devices that compose the network and can rearrange them in different configurations. Its core idea is to explore simple (linear) conditions of this inflated network that ultimately translate to polynomial inequalities on the observable probabilities. Navascués and Wolfe [58] proved the existence of a hierarchy of inflations that asymptotically converges to the classical set of correlations of any network, and a test of compatibility of a given level of this hierarchy can be done via linear programming [63]. However, for each level n of this hierarchy, the memory resources required are superexponential in n . Here we show how we can use e-separation to drastically decrease the number of variables for each compatibility test and prove the Bonet PR-box incompatibility.

In parallel to the concept of inflation, the concept of e-separation, short for extended d-separation, was introduced in Ref. [48]. The e-separation criterion is a graphical

method that allows us to derive, instead of equality constraints like in d-separation, inequality constraints on the probabilities. This criterion is related to the idea of splitting nodes in a graph that allows considering another equivalent situation where we can proceed to use the inflation technique. We can define the e-separation criterion in terms of the d-separation criterion as follows.

Definition 2: Let \mathbf{A} , \mathbf{B} , \mathbf{C} , and \mathbf{D} be disjoint sets of variables in DAG \mathcal{G} . We say that the variables \mathbf{A} and \mathbf{B} are e-separated given \mathbf{C} after deletion of \mathbf{D} only if \mathbf{A} and \mathbf{B} are d-separated by \mathbf{C} in DAG \mathcal{G}^* , which is the resulting DAG after removal of all the vertices in \mathbf{D} from \mathcal{G} .

If $\mathbf{D} = \emptyset$, e-separation recovers the notion of d-separation.

In practice, we can see e-separation in terms of a node-splitting operation and independence as follows. Given a graph \mathcal{G} and a vertex \mathbf{D} in the graph, the node-splitting operation returns a new graph $\mathcal{G}^\#$ in which \mathbf{D} is split into two vertices. One of the vertices, denoted \mathbf{D} , maintains all its causal parents in the original graph \mathcal{G} , thus having the same distribution as \mathbf{D} in \mathcal{G} , but none of its outgoing edges. The other one, labeled $\mathbf{D}^\#$, will instead inherit all of the outgoing edges of \mathbf{D} in the original graph, but none of its incoming ones. In Ref. [48] it was proved that \mathbf{A} and \mathbf{B} are e-separated given \mathbf{C} after deletion of \mathbf{D} if and only if \mathbf{A} and \mathbf{B} are d-separated by \mathbf{C} and $\mathbf{D}^\#$ in $\mathcal{G}^\#$, i.e., $q(a, b|c, d^\#) = q(a|c, d^\#)q(b|c, d^\#)$. Therefore, we can say that e-separation is equivalent to the set of conditions

$$\exists q(a, b, d|c, d^\#) \text{ conditional probability distribution} \quad (28a)$$

$$\text{such that } q(a, b, d|c, d^\# = d) = p(a, b, d|c), \quad (28b)$$

$$q(a, b|c, d^\#) = q(a|c, d^\#)q(b|c, d^\#). \quad (28c)$$

Since the conditions are only bound on the existence of the variables $q(a, b, d|c, d^\#)$, one could perform variable

elimination to consider only the conditions imposed to the $p(a, b, d|c)$ variables.

In the Evans scenario, this would be the case if we identify the nodes $\mathbf{A} = A$, $\mathbf{B} = C$, $\mathbf{C} = \emptyset$, and $\mathbf{D} = B$ (see Fig. 4). We can see that e-separation allows us to derive constraints on our distribution p over the original variables by exploring the independence relations of distribution q over the variables of the DAG after performing a node-splitting operation. We can use a similar idea for the inflation technique. Instead of the observable d-separation relations of q , we could go a step further and explore any condition imposed by the DAG. In particular, the nontrivial conditions imposed on q by inflation will yield nontrivial inflation-type conditions on p .

After performing the node-splitting operation, we use the inflation technique in the resulting DAG. For simplicity, let us show an example with second-order inflation.

Considering two independent copies of each exogenous variable λ_i , μ_i , $B_i^\#$ for $i = 1, 2$ and $A_i = A(\lambda_i, B_i^\#)$, $C_j = C(\mu_j, B_j^\#)$, and $B_{ij} = B(\lambda_i, \mu_j)$, give us a joint distribution on all observed variables, since the $B^\#$ variables are observable, we can interpret them as inputs. Suppose that q is compatible with the DAG on the right of Fig. 4, i.e.,

$$q(a, b, c|b^\#) = \sum_{\lambda, \mu} q(\lambda)q(\mu)q(a|\lambda, b^\#)q(b|\lambda, \mu)q(c|\mu, b^\#); \quad (29)$$

then there exists a joint distribution on the inflated DAG respecting

$$\begin{aligned} q'(a_1, a_2, b_{11}, b_{12}, b_{21}, b_{22}, c_1, c_2|b_1^\#, b_2^\#) \\ = \sum_{\lambda_1, \mu_1, \lambda_2, \mu_2} q(\mu_1)q(\lambda_1)q(\mu_2)q(\lambda_2)q(a_1|\lambda_1 b_1^\#)q \\ \times (a_2|\lambda_2 b_2^\#) \prod_{i,j=1,2} q(b_{ij}|\lambda_i \mu_j)q(c_1|\mu_1 b_1^\#)q(c_2|\mu_2 b_2^\#). \end{aligned} \quad (30)$$

Although we cannot test Eq. (30) directly, the equation imposes linear constraints on distribution q . Switching

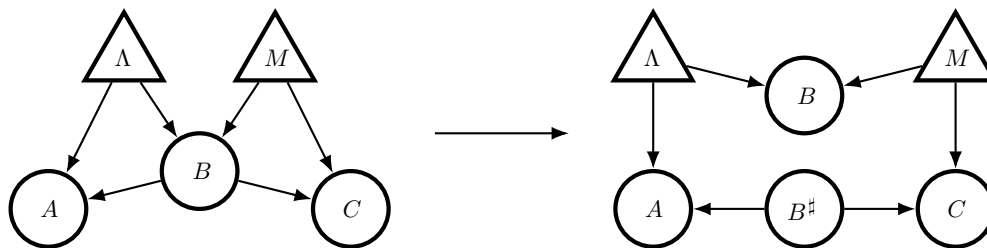


FIG. 4. Node-splitting operation. In the node-splitting operation, DAG \mathcal{G} on the left, in this case representing the Evans scenario, is associated with DAG $\mathcal{G}^\#$ on the right, where one of the nodes, node B , has been split into two nodes B and $B^\#$, respectively retaining the incoming and outgoing edges of the original node. DAG $\mathcal{G}^\#$ is useful to visualize the e-separation condition between nodes A and C deleting B , which here corresponds to the d-separation condition $A \perp_d C | B^\#$.

the labels $(\mu_1 \leftrightarrow \mu_2, \lambda_1 \leftrightarrow \lambda_2, a_1 \leftrightarrow a_2, c_1 \leftrightarrow c_2, b_{11} \leftrightarrow b_{22}, b_{12} \leftrightarrow b_{21}, b_1^\# \leftrightarrow b_2^\#)$ leaves Eq. (30) unchanged. This implies that

$$\begin{aligned} q'(a_1, a_2, b_{11}, b_{12}, b_{21}, b_{22}, c_1, c_2 | b_1^\# b_2^\#) \\ = q'(a_2, a_1, b_{22}, b_{21}, b_{12}, b_{11}, c_2, c_1 | b_2^\# b_1^\#). \end{aligned} \quad (31)$$

We may also consider the same expression when $b_1^\# = b_2^\# = b^\#$; switching the labels $(\lambda_1 \leftrightarrow \lambda_2, a_1 \leftrightarrow a_2, b_{11} \leftrightarrow b_{21}, b_{12} \leftrightarrow b_{22})$ or $(\mu_1 \leftrightarrow \mu_2, c_1 \leftrightarrow c_2, b_{11} \leftrightarrow b_{12}, b_{21} \leftrightarrow b_{22})$ yields

$$\begin{aligned} q'(a_1, a_2, b_{11}, b_{12}, b_{21}, b_{22}, c_1, c_2 | b_1^\# = b_2^\# = b^\#) \\ = q'(a_2, a_1, b_{21}, b_{22}, b_{11}, b_{12}, c_1, c_2 | b_1^\# = b_2^\# = b^\#) \\ = q'(a_1, a_2, b_{12}, b_{11}, b_{22}, b_{21}, c_2, c_1 | b_1^\# = b_2^\# = b^\#), \end{aligned} \quad (32)$$

and it has to satisfy some trivial conditions such as normalization, non-negativity, and the fact that outputs a_i, c_i are independent of inputs $b_{j \neq i}^\#$. Finally, condition (30) tells us that q' recovers q via marginalization, but only some entries of q are available since we, initially, only have p . This yields a set of linear conditions

$$\begin{aligned} q'(\{a_i, b_{ii}, c_i | b_i^\# = b_{ii}\}) &= \prod_i q(a_i, b_{ii}, c_i | b_i^\# = b_{ii}) \\ &= \prod_i p(a_i, b_{ii}, c_i). \end{aligned} \quad (33)$$

An analogous procedure also follows for the general case where, in the n th order, we would consider n independent copies of Λ , M , and $B^\#$ and n^2 copies of $A(\lambda, B^\#)$, $B(\lambda, \mu)$, and $C(\mu, B^\#)$. We need to consider all the relabelings that leave the Markov model of the resulting network invariant and recover n -degree polynomials over the diagonal distribution $q'_n(\{A_i, B_{ii}, C_i | B_i^\# = B_{ii}\})$. For the Bonet PR box (26), we are able to find incompatibility for $n = 3$. It is also important to note that e-separation + second-order inflation yields, for our case where $|B| = |C| = 2$ and $|A| = 3$, a linear program (LP) with 2304 variables, and for third level, yields 884736 variables, while the hierarchy detailed in Ref. [58] yields 20736 variables for second-order inflation and 23887872 variables for third-order inflation, which would be quite challenging even for a LP.

We can translate the unfeasible status of our certification in terms of a witness via Farkas' lemma [63], which states that either the linear system $Ax = b, x \geq 0$, has a solution or $A^T y \leq 0$ has a solution with $b^T y < 0$. Thus, if our linear programming certification has no solution for a given p , there exists a solution y_{op} such that the symbolic

expression

$$b^T y_{op} \geq 0 \quad (34)$$

can be understood as a causal inequality, where $b = b(p)$ is a vector that has entries of all monomials of degree at most n due to Eq. (33). Applying this procedure directly on our Bonet PR box with $p_A(0) = p_A(2) = 10/21$ and $p_A(1) = 1/21$ we are able to retrieve such a witness in terms of cubic monomials of $p(a, b, c)$. For simplicity, we could make the mild assumption that $p(c = f(a, b) + b) = 1$, which can be guaranteed under classical models and is true for our candidate distribution (26), to write down the inequality as

$$W(p) \leq 1/3, \quad (35)$$

where $W(p)$ is a cubic polynomial that we explicitly show in Appendix B. Our candidate distribution reaches a violation of $\beta_{PR} := W(p_{PR}) - 1/3 \approx 3.2 \times 10^{-3}$ and the inequality gives us a white noise visibility of $v = 0.9984$. We were not able to find any quantum violations of this witness.

VI. DERIVING THEORY-INDEPENDENT CONSTRAINTS

So far, we have focused on the derivation of constraints valid for a classical description of Evans causal structure. In some cases, however, it might be interesting to have constraints valid for generalized probabilistic theories, that is, constraints reflecting the topology of the causal network rather than the nature (classical, quantum, or even postquantum) of the sources. These theory-independent constraints can be seen as genuine witnesses of the topology of the network and have already been derived for certain classes of causal networks [10,40,64], most prominently the triangle network [65,66]. In the following, we propose a general route for deriving such theory-independent constraints for the Evans causal structure and derive one specific inequality to demonstrate the method.

A crucial difference between a classical and a general probabilistic description of a causal network stems from the fact that the local changes performed by any part in the network cannot be arbitrary as, for instance, they may possess resources that cannot be perfectly cloned. That is, differently from the classical case, we can only consider non-fan-out inflations, as we cannot broadcast information that comes from nonclassical latent variables. An inflationary fan-out is a latent node that has two or more children that are copies of each other. Graphically, this would be the creation of another outgoing arrow from a latent node; see Fig. 6. Thus, in non-fan-out inflations, we cannot change the numbers of outgoing edges of the latent variables. It

has been noted in Ref. [42] that inflations that do not possess inflationary fan-outs will only yield conditions that are GPT valid.

Recently, Gisin *et al.* [65] showed a relationship between non-fan-out inflations and the assumptions of no signaling and independence of the sources in networks. Namely, one part may potentially signal to the others by locally modifying the structure of the network by adding sources and measurement devices. Therefore, conditions implied by non-fan-out inflations can be interpreted as restrictions that will arise for no-signaling theories in networks.

Looking from this perspective, we can see that in the node-splitting operation, discussed in the last section, the inward arrows to A and C do not change. Therefore, we can view this operation as a local modification in the network topology performed by B . Thus, e-separation should also follow for all no-signaling behaviors. In fact, it was shown in Refs. [48,67] that e-separation alone can impose constraints on the GPT behaviors of the network. However, we can recursively employ these local operations in our network. Namely, after some part performs a local operation, we can end up with a new network in which the new parts are independent and are also allowed to locally change the network structure. Note that in the last section we used e-separation to derive constraints on the classical set of correlations; these constraints are violated by the PR box due to the explicit use of inflationary fan-outs that only follow for classical latent variables.

Inspired by the hexagon inflation in Ref. [65], Bob can set up this new experiment such that the laboratory of Bob[#] is very far away and Bob[#] may adopt a similar strategy,

again changing the network topology, and this should be oblivious to Alice, Bob, and Charlie (see Fig. 5). Note that if we were to consider two independent copies of node $B^\#$, which we could, we would retrieve the hexagon by interpreting the $B^\#$ nodes as observable shared randomness between the parts. In our case, the d-separation and marginal independence imply nonlinear constraints involving nonobservable terms, posing a more challenging nonlinear quantifier elimination procedure. Fortunately, we do not need to consider independent copies of the $B^\#$ variable, since it is an observable variable and therefore we can assume it to be classical regardless of the resources the parts share. This means that the information sent to A and C can be perfectly cloned; thus, we may broadcast it throughout the network. With these remarks, the conditions implied are

$$q(a, a', b = b', c, c' | b^\# = b) = p(a, b, c)p(a', b' = b, c'), \tag{36}$$

as well as

$$q(b, b' | b^\#) = q(b, b') \quad \text{for all } b^\#,$$

and normalization conditions. These equality constraints together with the non-negativity conditions will form a set of linear inequalities on the variables q . Since the quadratic terms of p may be treated like symbolic constants, we may perform Fourier-Motzkin elimination to arrive at quadratic inequalities that should follow for all no-signaling behaviors.

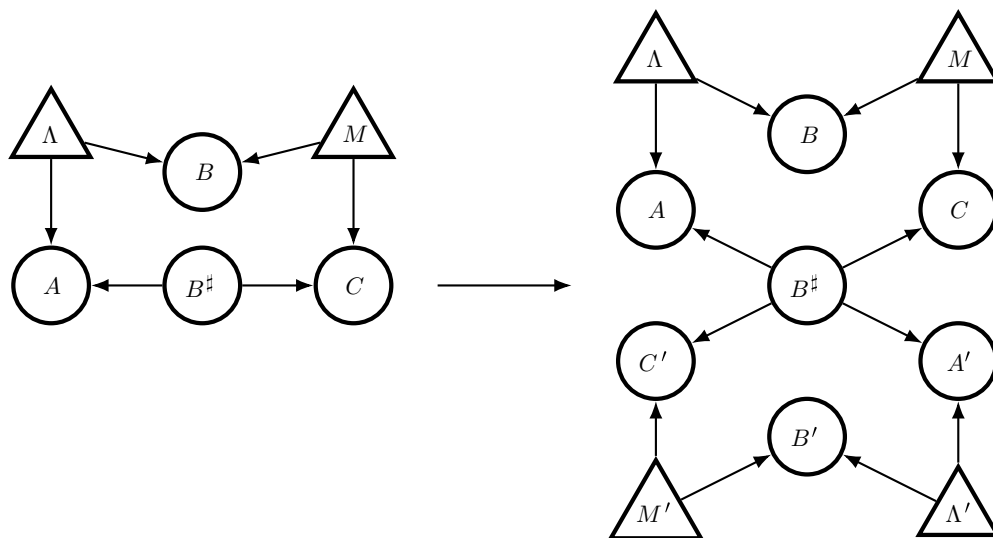


FIG. 5. Inflation of the Evans scenario. In order to derive the topological constraints of the scenario on the left, we consider the non-fan-out inflation on the right. Importantly, the operations needed to bring the original Evans scenario to the node splitted DAG and then to the inflated causal structure follow for all generalized probabilistic theories and can be performed locally by the parts that comprise the network.

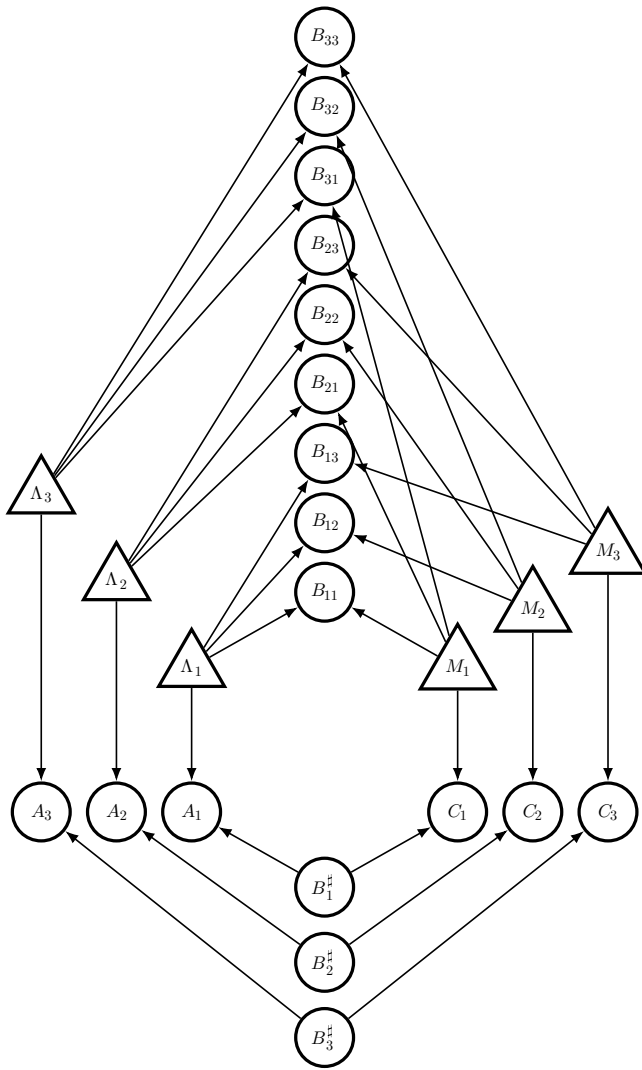


FIG. 6. Inflated Evans scenario where Λ_i and M_i are copies of the original sources Λ and M , and A_i , $B_{i,j}$, C_i , and $B_i^\#$ are copies of the original observable variables A , B , C , and $B^\#$. The indices of A , B , and C indicate the latent dependency of each observable copy variable.

The procedure outlined above is quite general and could be used to derive all constraints following a given inflation. However, it can be very costly and out of computational reach even for seemingly simple scenarios such as the Evans scenario. To circumvent that, similarly to the approach discussed in the last section, given a behavior from the Evans scenario, we may employ a LP to detect incompatibility between the distribution and the set of inequalities. That is, given a distribution $p(a, b, c)$, we solve

$$\max_q \quad 1 \quad (37a)$$

$$\text{such that } \begin{aligned} q(a, a', b = b', c, c' | b^\# = b) &= p \\ (a, b, c)p(a', b' = b, c') &, \end{aligned} \quad (37b)$$

$$q(b, b' | b^\#) = q(b, b') \quad \text{for all } b^\#, \quad (37c)$$

$$\sum q(a, a', b, b', c, c' | b^\#) = 1 \quad \text{for all } b^\#, \quad (37d)$$

$$q(a, a', b, b', c, c' | b^\#) \geq 0 \quad \text{for all } a, a', b, b', c, c', b^\#. \quad (37e)$$

If this LP is feasible, behavior $p(a, b, c)$ is compatible with the constraints. Otherwise, it is not. If we have a nonfeasible distribution, we can then rely on the dual of LP (37) to easily find the corresponding witness (a nonlinear inequality) detecting it.

Using this idea, we can show that distribution $p_{\text{unfeasible}}(a, b, c)$ given by $p(0, 0, 0) = p(1, 0, 1) = 1/2$ is incompatible with the Evans topology, as it returns an unfeasible LP. From the dual problem, we can obtain a witness for the incompatibility, which reads

$$\begin{aligned} & p(0, 0, 0)^2 + p(0, 0, 1)p(0, 0, 0) + p(0, 0, 1)^2 \\ & + 4p(1, 0, 1)p(0, 0, 0) + p(1, 0, 0)^2 \\ & + p(1, 0, 1)p(1, 0, 0) + p(1, 0, 1)^2 \leq 1. \end{aligned} \quad (38)$$

This theory-independent polynomial inequality is violated by the distribution $p(0, 0, 0) = p(1, 0, 1) = 1/2$ as the right-hand side of it reaches the value $3/2$. Furthermore, this constraint turns out to be quite resistant to noise, as can be seen by considering the mixed distribution $vp_{\text{unfeasible}} + (1-v)p_W$ (p_W being the white noise distribution where all binary outcomes have the same $1/8$ probability). The threshold visibility for the inequality violation is $v_{\text{crit}} \approx 0.7398$.

VII. DISCUSSION

Bell's theorem is the most stringent notion of nonclassicality since it only relies on causal assumptions about the experimental apparatus, but remains agnostic of any internal mechanisms or physical details of the measurement and state preparation devices. Given its foundational as well as applied consequences, generalizing such results to causal structures of increasing complexity and with different topologies is a timely and promising research direction.

Within this context, the contribution of this paper is to analyze a simple yet elusive causal structure. It is known that there are three different classes of causal structures with three observable variables that can potentially give rise to a mismatch between classical and quantum predictions. Two of those are indeed known to support quantum nonclassicality and have been the focus of much of the literature on the topic recently. On the one hand, we have the instrumental causal structure [31,45], a scenario of crucial relevance for estimating causal influences [68] and

that, in particular, brings to light the relevance of considering interventions, rather than pure passive observations, in order to reveal the nonclassicality of a given quantum process [32,33]. On the other hand, we have the triangle scenario [15,18] that, differently from Bell’s theorem, has the central feature of not requiring locally incompatible measurement observables and solely resorting to the independence of sources for leading to nonclassical behaviors [29]. Here we focus on the third of these causal structures, the Evans scenario [34], a sort of blend between the instrumental and triangle structures since it involves independent sources and no external inputs but does presume direct causal influences between the measurement outcomes.

Being akin to the entanglement swapping experiment [52], it is natural to expect that the correlations obtained by considering a Bell state measurement performed on shared Bell states should lead to nonclassical correlations. Quite the opposite, we proved that such an experiment does have a classical causal model for any dimension. So, in order to prove nonclassicality, we explored different venues, establishing the connections between the Evans scenario and the instrumental and bilocality scenarios. In particular, we proved that nonclassicality in the bilocality structure is a necessary but not sufficient condition for nonclassicality in the Evans case. Unfortunately, however, this bridge with well-known scenarios proved insufficient to find nonclassicality in the Evans structure, motivating us to search for alternative methods to study the set of correlations entailed by it.

With that in mind, we discussed the features of the Gurobi optimizer [69], capable of handling nonconvex quadratic optimization problems and being perfectly fitted to analyze the Evans causal structure. With the Gurobi optimizer, we could not only numerically prove that the analogous of a PR box in the instrumental scenario does lead to nonclassical correlations in the Evans scenario, but also recover the numerical witness for it. Going beyond this numerical study, we combined the inflation [42] and e-separation [48] techniques to derive polynomial Bell-like inequalities that can be violated, an unambiguous proof that the Evans causal structure can also support nonclassical correlations, even though they are of a postquantum nature. Finally, we showed how to derive theory-independent constraints for the Evans scenario, that is, inequalities that reflect the topology of the network, similar to what has been done in the triangle scenario [65].

The natural next step is to resolve whether quantum nonclassical behavior is or is not possible in the Evans scenario. In particular, if one could prove that all quantum correlations do have a classical explanation, this could lead to a novel situation. Note that all principles for quantum correlations focus not on a given causal structure but rather on particular Bell inequalities entailed by it. If one could prove that a particular causal structure does not have any quantum advantage (but has a postquantum

one, as we prove here), this would open a totally new venue to understand quantum correlations and their underlying principles. Finally, it is worth remarking that here we focused only on the observed distribution $p(a, b, c)$. However, as shown recently for the instrumental scenario [32], interventions could be a possible way to unveil the nonclassicality, that is, resorting to interventional data of the form $p(a, c|do(b))$, where $do(b)$ refers to interventions in the central node of the network. Either way, the resolution of this problem will certainly be of relevance to the larger question of understanding what are the crucial ingredients of the emergence of nonclassical behavior in causal structures of growing size and complexity.

ACKNOWLEDGMENTS

We thank Elie Wolfe for fruitful discussions. This work is supported by the Serrapilheira Institute (Grant No. Serra-1708-15763), by the Simons Foundation (Grant No. 884966, AF), the Brazilian National Council for Scientific and Technological Development (CNPq) via the National Institute for Science and Technology on Quantum Information (INCT-IQ) Grant No. and 307295/2020-6, and the São Paulo Research Foundation FAPESP (Grants No. 2018/07258-7 and No. 2022/03792-4).

APPENDIX A: CLASSICAL SIMULATION OF THE PR BOX WITH SYMMETRIC INSTRUMENT

We show how the correlations of PR box (26) between Bob and Charlie with uniform distribution of Alice can be simulated classically. The distribution under scrutiny is

$$p(a, b, c) = p(a)p(bc|a) = \begin{cases} \frac{1}{6} & \text{if } c = f(a, b) + b, \\ 0 & \text{otherwise.} \end{cases} \quad (\text{A1})$$

It is enough to take $\lambda = \lambda_0\lambda_1$, $\lambda_i \in \{0, 1, 2\}$, and $\mu = \mu_0\mu_1$, $\mu_i \in \{0, 1\}$, with distributions

$$p(\gamma) = ([01] + [11] + [20] + [22])/4 \\ \text{and } p(\alpha) = [10]/3 + 2[01]/3, \quad (\text{A2})$$

where $[a_0a_1] = \delta_{\lambda_0, a_0}\delta_{\lambda_1, a_1}$ and $[c_0c_1] = \delta_{\mu_0, c_0}\delta_{\mu_1, c_1}$. Define deterministic response functions for Alice and Charlie as

$$p(a|\lambda, b) = \delta_{a, \lambda_b}, \quad p(c|\mu, b) = \delta_{c, \mu_b}. \quad (\text{A3})$$

Bob also has a deterministic local response function

$$p(b|\lambda\mu) = \begin{cases} 1 & \text{if } \mu_b = f(\lambda_b, b) + b, \\ 0 & \text{otherwise.} \end{cases} \quad (\text{A4})$$

Because of our choice of strategies, this response function is well defined, i.e., the expression $\mu_b = f(\lambda_b, b) + b$

has a unique solution b for every strategy $\lambda\mu$ given to Bob. In our case, since b and μ_i are bits, we can write $b = \mu_b + f(\lambda_b, b)$, by summing $b + \mu_b$ on both sides. The deterministic response can be rewritten as

$$p(b|\lambda\mu) = \delta_{b, \mu_b + f(\lambda_b, b)}. \quad (\text{A5})$$

Summing up, this implies that

$$\begin{aligned} p(a, b, c) &= \sum_{\lambda, \mu} p(\lambda)p(\mu)\delta_{a, \lambda_b}\delta_{b, \mu_b + f(\lambda_b, b)}\delta_{c, \mu_b} \\ &= \delta_{b, c + f(a, b)} \left(\sum_{\lambda, \mu} p(\lambda_b = a)p(\mu_b = c) \right), \quad (\text{A6}) \end{aligned}$$

and we can check that, for each case where $b = c + f(a, b)$, the sum above is $1/6$, and thus this classical model recovers the distribution $p(a, b, c)$.

APPENDIX B: INEQUALITY DERIVED FROM INFLATION WITH E-SEPARATION

Here we explicitly show the inequality derived using third-order inflation after the node-splitting operation in

the Evans scenario. The inequality is obtained via the dual solution of the linear program detailed in Sec. V. We consider the inflation shown in Fig. 6, where the compatibility test can be cast as a linear program on the joint probability distribution

$$q(a_1, a_2, a_3, b_{11}, b_{12}, b_{13}, b_{21}, b_{22}, b_{23}, b_{31}, b_{32}, b_{33}, c_1, c_2, c_3 | b_1^\#, b_2^\#, b_3^\#) \quad (\text{B1})$$

that must respect non-negativity and normalization. Defining

$$\begin{aligned} q_3(a_1, a_2, a_3, b_{11}, b_{22}, b_{33}, c_1, c_2, c_3 | b_1^\#, b_2^\#, b_3^\#) \\ = \sum_{b_i \neq j} q(a_1, \dots, b_{11}, \dots, c_3 | b_1^\#, b_2^\#, b_3^\#), \quad (\text{B2}) \end{aligned}$$

the factorization constraints becomes

$$\begin{aligned} q_3(a_1, a_2, a_3, b_{11}, b_{22}, b_{33}, c_1, c_2, c_3 | (b_1^\#, b_2^\#, b_3^\#)) \\ = (b_{11}, b_{22}, b_{33}) = \prod_{i=1,2,3} p(a_i, b_{ii}, c_i), \quad (\text{B3}) \end{aligned}$$

where $p(a, b, c)$ is the probability distribution under scrutiny. Ultimately, we need to impose symmetry constraints implied by the global Markov model of the inflated network; these conditions will be simply equality constraints on the entries of q . This can be done in analogy to Eqs. (31) and (32) for when $(b_1^\#, b_2^\# = b_3^\#)$, $(b_1^\# = b_2^\#, b_3^\#)$, and $(b_1^\# = b_2^\# = b_3^\#)$. Because of the factorization conditions imposed in the diagonal distribution q_3 , the obtained inequality contains the corresponding cubic elements and is given by

$$\begin{aligned} & p_{ABC}(0, 0, 0)^3/3 + p_{ABC}(0, 0, 0)^2 p_{ABC}(0, 1, 1) + p_{ABC}(0, 0, 0)^2 p_{ABC}(1, 0, 0) \\ & + p_{ABC}(0, 0, 0)^2 p_{ABC}(1, 1, 0) + p_{ABC}(0, 0, 0)^2 p_{ABC}(2, 0, 1) + p_{ABC}(0, 0, 0)^2 p_{ABC}(2, 1, 1) \\ & + p_{ABC}(0, 0, 0) p_{ABC}(0, 1, 1)^2 + p_{ABC}(0, 0, 0) p_{ABC}(2, 1, 1)^2 \\ & + 2p_{ABC}(0, 0, 0) p_{ABC}(0, 1, 1) p_{ABC}(1, 0, 0) + p_{ABC}(0, 1, 1)^3/3 \\ & + p_{ABC}(0, 1, 1)^2 p_{ABC}(1, 0, 0) + p_{ABC}(0, 1, 1)^2 p_{ABC}(1, 1, 0) \\ & + 2p_{ABC}(0, 0, 0) p_{ABC}(0, 1, 1) p_{ABC}(2, 0, 1) + 2p_{ABC}(0, 0, 0) p_{ABC}(0, 1, 1) p_{ABC}(1, 1, 0) \\ & + p_{ABC}(2, 0, 1)^2 p_{ABC}(2, 1, 1) + 2p_{ABC}(0, 0, 0) p_{ABC}(1, 0, 0) p_{ABC}(1, 1, 0) \\ & + 2p_{ABC}(0, 0, 0) p_{ABC}(1, 0, 0) p_{ABC}(2, 0, 1) + p_{ABC}(2, 0, 1) p_{ABC}(2, 1, 1)^2 \\ & + 2p_{ABC}(0, 0, 0) p_{ABC}(1, 0, 0) p_{ABC}(2, 1, 1) + p_{ABC}(0, 0, 0) p_{ABC}(1, 1, 0)^2 \\ & + p_{ABC}(0, 0, 0) p_{ABC}(1, 0, 0)^2 + 2p_{ABC}(0, 0, 0) p_{ABC}(1, 1, 0) p_{ABC}(2, 0, 1) \\ & + 2p_{ABC}(0, 0, 0) p_{ABC}(1, 1, 0) p_{ABC}(2, 1, 1) - 5p_{ABC}(0, 0, 0) p_{ABC}(2, 0, 1)^2/3 \\ & + p_{ABC}(0, 1, 1)^2 p_{ABC}(2, 0, 1) + p_{ABC}(0, 1, 1)^2 p_{ABC}(2, 1, 1) \\ & + 2p_{ABC}(0, 0, 0) p_{ABC}(2, 0, 1) p_{ABC}(2, 1, 1) + p_{ABC}(0, 1, 1) p_{ABC}(1, 0, 0)^2 \\ & + 2p_{ABC}(0, 1, 1) p_{ABC}(1, 0, 0) p_{ABC}(1, 1, 0) + p_{ABC}(1, 1, 0) p_{ABC}(2, 1, 1)^2 \end{aligned}$$

$$\begin{aligned}
 &+ 2p_{ABC}(0, 1, 1)p_{ABC}(1, 0, 0)p_{ABC}(2, 0, 1) + 2p_{ABC}(0, 1, 1)p_{ABC}(1, 0, 0)p_{ABC}(2, 1, 1) \\
 &- p_{ABC}(2, 0, 1)^3 + p_{ABC}(0, 1, 1)p_{ABC}(1, 1, 0)^2 + 14p_{ABC}(0, 1, 1)p_{ABC}(1, 1, 0)p_{ABC}(2, 0, 1)/3 \\
 &+ 2p_{ABC}(0, 0, 0)p_{ABC}(0, 1, 1)p_{ABC}(2, 1, 1) + 2p_{ABC}(0, 1, 1)p_{ABC}(1, 1, 0)p_{ABC}(2, 1, 1) \\
 &+ p_{ABC}(0, 1, 1)p_{ABC}(2, 0, 1)^2 + 2p_{ABC}(0, 1, 1)p_{ABC}(2, 0, 1)p_{ABC}(2, 1, 1) \\
 &+ p_{ABC}(0, 1, 1)p_{ABC}(2, 1, 1)^2 + p_{ABC}(1, 0, 0)^3/3 + p_{ABC}(1, 0, 0)p_{ABC}(2, 1, 1)^2 \\
 &+ p_{ABC}(1, 1, 0)^3/3 + p_{ABC}(1, 0, 0)^2p_{ABC}(1, 1, 0) + p_{ABC}(1, 0, 0)^2p_{ABC}(2, 0, 1) \\
 &+ p_{ABC}(1, 0, 0)^2p_{ABC}(2, 1, 1) + p_{ABC}(2, 1, 1)^3/3 + p_{ABC}(1, 0, 0)p_{ABC}(1, 1, 0)^2 \\
 &+ 2p_{ABC}(1, 0, 0)p_{ABC}(1, 1, 0)p_{ABC}(2, 0, 1) + 14p_{ABC}(1, 1, 0)p_{ABC}(2, 0, 1)p_{ABC}(2, 1, 1)/3 \\
 &+ 2p_{ABC}(1, 0, 0)p_{ABC}(1, 1, 0)p_{ABC}(2, 1, 1) - 5p_{ABC}(1, 0, 0)p_{ABC}(2, 0, 1)^2/3 \\
 &+ 2p_{ABC}(1, 0, 0)p_{ABC}(2, 0, 1)p_{ABC}(2, 1, 1) + p_{ABC}(1, 1, 0)^2p_{ABC}(2, 0, 1) \\
 &+ p_{ABC}(1, 1, 0)^2p_{ABC}(2, 1, 1) + p_{ABC}(1, 1, 0)p_{ABC}(2, 0, 1)^2 \leq 1/3.
 \end{aligned} \tag{B4}$$

For simplicity, we have assumed that $p(c = f(a, b) + b) = 1$ to keep only the nonzero terms of the inequality for our candidate distribution. This condition is also not too restrictive as it can be guaranteed under classical models, for example, $p(0, 0, 0) = 1$ trivially satisfies this constraint. If we choose $p_A(0) = p_A(2) = 10/21$ and $p_A(1) = 1/21$, the distribution violates inequality (B4) by $\beta_{PR} \approx 3.2 \times 10^{-3}$ and has a very low resistance when combined with the uniform distribution with visibility $v \approx 0.9984$.

[1] J. S. Bell, On the Einstein Podolsky Rosen paradox, *Phys. Phys. Fiz.* **1**, 195 (1964).
 [2] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Bell nonlocality, *Rev. Mod. Phys.* **86**, 419 (2014).
 [3] S. Pironio, V. Scarani, and T. Vidick, Focus on device independent quantum information, *New J. Phys.* **18**, 100202 (2016).
 [4] J. F. Fitzsimons, J. A. Jones, and V. Vedral, Quantum correlations which imply causation, *Sci. Rep.* **5**, 18281 (2015).
 [5] K. Ried, M. Agnew, L. Vermeyden, D. Janzing, R. W. Spekkens, and K. J. Resch, A quantum advantage for inferring causal structure, *Nat. Phys.* **11**, 414 (2015).
 [6] T. Fritz, Beyond Bell’s theorem II: Scenarios with arbitrary causal structure, *Commun. Math. Phys.* **341**, 391 (2016).
 [7] R. Chaves, C. Majenz, and D. Gross, Information-theoretic implications of quantum causal structures, *Nat. Commun.* **6**, 1 (2015).
 [8] F. Costa and S. Shrapnel, Quantum causal modelling, *New J. Phys.* **18**, 063032 (2016).
 [9] 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).

[10] J. Åberg, R. Nery, C. Duarte, and R. Chaves, Semidefinite Tests for Quantum Network Topologies, *Phys. Rev. Lett.* **125**, 110505 (2020).
 [11] E. Wolfe, A. Pozas-Kerstjens, M. Grinberg, D. Rosset, A. Acín, and M. Navascués, Quantum Inflation: A General Approach to Quantum Causal Compatibility, *Phys. Rev. X* **11**, 021043 (2021).
 [12] L. T. Ligthart, M. Gachechiladze, and D. Gross, A convergent inflation hierarchy for quantum causal structures, (2021), arXiv preprint [ArXiv:2110.14659](https://arxiv.org/abs/2110.14659).
 [13] A. Tavakoli, A. Pozas-Kerstjens, M. X. Luo, and M.-O. Renou, Bell nonlocality in networks, *Reports on Progress in Physics* (2021).
 [14] C. Branciard, N. Gisin, and S. Pironio, Characterizing the Nonlocal Correlations Created via Entanglement Swapping, *Phys. Rev. Lett.* **104**, 170401 (2010).
 [15] T. Fritz, Beyond Bell’s theorem: Correlation scenarios, *New J. Phys.* **14**, 103001 (2012).
 [16] A. Tavakoli, P. Skrzypczyk, D. Cavalcanti, and A. Acín, Nonlocal correlations in the star-network configuration, *Phys. Rev. A* **90**, 062109 (2014).
 [17] F. Andreoli, G. Carvacho, L. Santodonato, R. Chaves, and F. Sciarrino, Maximal qubit violation of n-locality inequalities in a star-shaped quantum network, *New J. Phys.* **19**, 113020 (2017).
 [18] M.-O. Renou, E. Bäumer, S. Boreiri, N. Brunner, N. Gisin, and S. Beigi, Genuine Quantum Nonlocality in the Triangle Network, *Phys. Rev. Lett.* **123**, 140401 (2019).
 [19] A. Tavakoli, N. Gisin, and C. Branciard, Bilocal Bell Inequalities Violated by the Quantum Elegant Joint Measurement, *Phys. Rev. Lett.* **126**, 220401 (2021).
 [20] R. Chaves, G. Moreno, E. Polino, D. Poderini, I. Agresti, A. Suprano, M. R. Barros, G. Carvacho, E. Wolfe, and A. Canabarro, *et al.*, Causal Networks and Freedom of Choice in Bell’s Theorem, *PRX Quantum* **2**, 040323 (2021).
 [21] A. Pozas-Kerstjens, N. Gisin, and A. Tavakoli, Full Network Nonlocality, *Phys. Rev. Lett.* **128**, 010403 (2022).

- [22] G. Carvacho, F. Andreoli, L. Santodonato, M. Bentivegna, R. Chaves, and F. Sciarrino, Experimental violation of local causality in a quantum network, *Nat. Commun.* **8**, 1 (2017).
- [23] D. J. Saunders, A. J. Bennet, C. Branciard, and G. J. Pryde, Experimental demonstration of nonbilocal quantum correlations, *Sci. Adv.* **3**, e1602743 (2017).
- [24] Q.-C. Sun, Y.-F. Jiang, B. Bai, W. Zhang, H. Li, X. Jiang, J. Zhang, L. You, X. Chen, and Z. Wang, *et al.*, Experimental demonstration of non-bilocality with truly independent sources and strict locality constraints, *Nat. Photon.* **13**, 687 (2019).
- [25] D. Poderini, I. Agresti, G. Marchese, E. Polino, T. Giordani, A. Suprano, M. Valeri, G. Milani, N. Spagnolo, and G. Carvacho, *et al.*, Experimental violation of n-locality in a star quantum network, *Nat. Commun.* **11**, 1 (2020).
- [26] G. Carvacho, R. Chaves, and F. Sciarrino, Perspective on experimental quantum causality, *EPL (Europhys. Lett.)* **125**, 30001 (2019).
- [27] H. Cao, M.-O. Renou, C. Zhang, G. Massé, X. Coiteux-Roy, B.-H. Liu, Y.-F. Huang, C.-F. Li, G.-C. Guo, and E. Wolfe, Experimental demonstration that no tripartite-nonlocal causal theory explains nature’s correlations, (2022), arXiv preprint [ArXiv:2201.12754](https://arxiv.org/abs/2201.12754).
- [28] A. Suprano, D. Poderini, E. Polino, I. Agresti, G. Carvacho, A. Canabarro, E. Wolfe, R. Chaves, and F. Sciarrino, Experimental genuine tripartite nonlocality in a quantum triangle network, (2022), arXiv preprint [ArXiv:2204.00388](https://arxiv.org/abs/2204.00388).
- [29] E. Polino, D. Poderini, G. Rodari, I. Agresti, A. Suprano, G. Carvacho, E. Wolfe, A. Canabarro, G. Moreno, and G. Milani, *et al.*, Experimental nonclassicality in a causal network without assuming freedom of choice, (2022), arXiv preprint [ArXiv:2210.07263](https://arxiv.org/abs/2210.07263).
- [30] J. B. Brask and R. Chaves, Bell scenarios with communication, *J. Phys. A: Math. Theor.* **50**, 094001 (2017).
- [31] R. Chaves, G. Carvacho, I. Agresti, V. Di Giulio, L. Aolita, S. Giacomini, and F. Sciarrino, Quantum violation of an instrumental test, *Nat. Phys.* **14**, 291 (2018).
- [32] M. Gachechiladze, N. Miklin, and R. Chaves, Quantifying Causal Influences in the Presence of a Quantum Common Cause, *Phys. Rev. Lett.* **125**, 230401 (2020).
- [33] I. Agresti, D. Poderini, B. Polacchi, N. Miklin, M. Gachechiladze, A. Suprano, E. Polino, G. Milani, G. Carvacho, and R. Chaves, *et al.*, Experimental test of quantum causal influences, *Sci. Adv.* **8**, eabm1515 (2022).
- [34] R. J. Evans, Graphs for margins of Bayesian networks, *Scand. J. Stat.* **43**, 625 (2016).
- [35] M. M. Ansanelli, *Looking for Quantum-Classical Gaps in Causal Structures* (Perimeter Institute, 2022).
- [36] L. D. Garcia, M. Stillman, and B. Sturmfels, Algebraic geometry of Bayesian networks, *J. Symb. Comput.* **39**, 331 (2005).
- [37] D. Geiger and C. Meek, Quantifier elimination for statistical problems, (2013), arXiv preprint [ArXiv:1301.6698](https://arxiv.org/abs/1301.6698).
- [38] R. Chaves, L. Luft, T. Maciel, D. Gross, D. Janzing, and B. Schölkopf, in *Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence* (2014), p. 112.
- [39] R. Chaves, Polynomial Bell Inequalities, *Phys. Rev. Lett.* **116**, 010402 (2016).
- [40] A. Kela, K. Von Prillwitz, J. Åberg, R. Chaves, and D. Gross, Semidefinite tests for latent causal structures, *IEEE Trans. Inf. Theory* **66**, 339 (2019).
- [41] C. M. Lee and R. W. Spekkens, Causal inference via algebraic geometry: Feasibility tests for functional causal structures with two binary observed variables, *J. Causal Inference* **5**, 20160013 (2017).
- [42] E. Wolfe, R. W. Spekkens, and T. Fritz, The inflation technique for causal inference with latent variables, *J. Causal Inference* **7**, 20170020 (2019).
- [43] A. Pozas-Kerstjens, R. Rabelo, Ł. Rudnicki, R. Chaves, D. Cavalcanti, M. Navascués, and A. Acín, Bounding the Sets of Classical and Quantum Correlations in Networks, *Phys. Rev. Lett.* **123**, 140503 (2019).
- [44] T. Kriváchy, Y. Cai, D. Cavalcanti, A. Tavakoli, N. Gisin, and N. Brunner, A neural network oracle for quantum nonlocality problems in networks, *npj Quantum Inf.* **6**, 1 (2020).
- [45] J. Pearl, in *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence* (1995), p. 435.
- [46] S. Popescu and D. Rohrlich, Quantum nonlocality as an axiom, *Found. Phys.* **24**, 379 (1994).
- [47] E. Wolfe, R. W. Spekkens, and T. Fritz, The inflation technique for causal inference with latent variables, *J. Causal Inference* **7**, 20170020 (2019).
- [48] R. J. Evans, in *2012 IEEE International Workshop on Machine Learning for Signal Processing* (2012), p. 1.
- [49] T. Verma and J. Pearl, in *Proc. 4th Workshop on Uncertainty in Artificial Intelligence (Minneapolis, MN and Mountain View, CA)* (1988).
- [50] G. Chiribella, G. M. D’Ariano, and P. Perinotti, Probabilistic theories with purification, *Phys. Rev. A* **81**, 062348 (2010).
- [51] R. J. Evans, Graphs for margins of Bayesian networks, *Scand. J. Stat.* **43**, 625 (2016).
- [52] M. Zukowski, A. Zeilinger, M. A. Horne, and A. K. Ekert, “Event-Ready-Detectors” Bell Experiment via Entanglement Swapping, *Phys. Rev. Lett.* **71**, 4287 (1993).
- [53] C. Branciard, N. Gisin, and S. Pironio, Characterizing the Nonlocal Correlations Created via Entanglement Swapping, *Phys. Rev. Lett.* **104**, 170401 (2010).
- [54] C. Branciard, D. Rosset, N. Gisin, and S. Pironio, Bilocal versus nonbilocal correlations in entanglement-swapping experiments, *Phys. Rev. A* **85**, 032119 (2012).
- [55] B. Bonet, Instrumentality tests revisited, (2013), arXiv preprint [ArXiv:1301.2258](https://arxiv.org/abs/1301.2258).
- [56] D. Rosset, N. Gisin, and E. Wolfe, Universal bound on the cardinality of local hidden variables in networks, *Quantum Inf. Comput.* **18**, 910 (2018).
- [57] J. B. Lasserre, Global optimization with polynomials and the problem of moments, *SIAM J. Optim.* **11**, 796 (2001).
- [58] M. Navascués and E. Wolfe, The inflation technique completely solves the causal compatibility problem, *J. Causal Inference* **8**, 70 (2020).
- [59] A. Fine, Hidden Variables, Joint Probability, and the Bell Inequalities, *Phys. Rev. Lett.* **48**, 291 (1982).
- [60] Gurobi Optimization, LLC, Gurobi Optimizer Reference Manual (2022).

- [61] The Gurobi optimizer supports constraints containing bilinear terms like $x \cdot y$. Although it does not directly support constraints containing more general multilinear terms, they can be modeled using a series of bilinear constraints, with the help of auxiliary variables. Such higher-degree polynomial problems will face numerical issues due to computational precision.
- [62] N. Miklin, M. Gachechiladze, G. Moreno, and R. Chaves, Causal inference with imperfect instrumental variables, *J. Causal Inference* **10**, 45 (2022).
- [63] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, 2006).
- [64] S. Beigi and M.-O. Renou, Covariance decomposition as a universal limit on correlations in networks, *IEEE Trans. Inf. Theory* **68**, 384 (2021).
- [65] N. Gisin, J.-D. Bancal, Y. Cai, P. Remy, A. Tavakoli, E. Zambrini Cruzeiro, S. Popescu, and N. Brunner, Constraints on nonlocality in networks from no-signaling and independence, *Nat. Commun.* **11**, 2378 (2020).
- [66] M.-O. Renou, Y. Wang, S. Boreiri, S. Beigi, N. Gisin, and N. Brunner, Limits on Correlations in Networks for Quantum and No-Signaling Resources, *Phys. Rev. Lett.* **123**, 070403 (2019).
- [67] N. Finkelstein, B. Zjawin, E. Wolfe, I. Shpitser, and R. W. Spekkens, in *UAI* (2021).
- [68] A. Balke and J. Pearl, Bounds on treatment effects from studies with imperfect compliance, *J. Am. Stat. Assoc.* **92**, 1171 (1997).
- [69] B. Bixby, The Gurobi optimizer, *Transp. Re-Search Part B* **41**, 159 (2007).