Polynomial Inequalities for Causal Inference with Hidden Variables

Elie Wolfe,[1,*] Robert W. Spekkens,[1,†] and Tobias Fritz[2,‡]

[1] *Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada, N2L 2Y5*

[2] *Max Planck Institute for Mathematics in the Sciences, Leipzig, Saxony, Germany, 04103*

(Dated: March 29, 2016)

The fundamental problem of causal inference is to infer from a given probability distribution over observed variables, what causal structures could have given rise to that distribution, possibly incorporating hidden variables. This kind of problem comes up in many fields. For example, Bell inequalities are criteria used in quantum theory to distinguish non-classical from classical correlations in very specific causal structures. Analogues of Bell inequalities which witness the incompatibility of a given distribution with a given causal structure are therefore highly sought after. We introduce a new class of such inequalities which apply to any causal structure (with hidden variables). This class contains all Bell inequalities as special cases.

Our technique consists of "inflating" the causal structure. We then translate weak constraints on the inflated structure into much stronger constraints on the original structure through "deflation" and quantifier elimination. One can derive either classical causal infeasibility criteria (i.e. that have quantum violations), or criteria which hold even in the context of general probability theories, depending on whether the inflation relies on cloning the values of the hidden variables. Concretely, we derive polynomial inequalities for the so-called Triangle scenario, and we show how the well-known Bell inequalities also follow from our method.Furthermore, our method can efficiently witness the incompatibility of a probability distribution with a given causal structure even absent explicit quantifier-free inequalities.

* ewolfe@perimeterinstitute.ca

† rspekkens@perimeterinstitute.ca

‡ tobias.fritz@mis.mpg.de

I. INTRODUCTION

Given some hypothesis of causal structure, it is desirable to determine the set of observable probability distributions compatible with the hypothesis. Causal infeasibility criteria are used in a wide variety of statistics application, from sussing out biological pathways to enabling artificially intelligent machine learning [1–4]. The foundational role of causal structure in quantum information theory has only recently been appreciated [5–8]: correlations feasible from classical causal structures are those probability distributions compatible with restricting the latent variables to be hidden ontic states. Quantum correlations, by contrast, are those which are uniquely realizable if the latent variables in the causal structure are allowed to be quantum states.

Tightly characterizing the set of observable probability distributions feasible from a causal structure is therefore physically critical, in order to recognize and exploit uniquely quantum distributions. Practical techniques for generically constraining causal compatibility include the use of conditional independence relations (easy) [1–4] and entropic inequalities (more advanced) [9–11]. These criteria, however, only rarely provide a tight characterization, and frequently fail to ascertain the non-classicality of quantum correlation.

Distinguishing quantum from classical correlations has historically been achieved through the use of Bell inequalities [12–15]. Bell inequalities however, are limited to causal scenarios involving *only one* latent common cause variable, i.e. Bell scenarios. A Bell scenario is special, however, in that it admits simple causal-feasibility characterization by means of a convex hull algorithm; this is because its associated set of feasible distributions is a convex polytope [13, 16]. The nonconvexity of distributions feasible from a general casual structure is explicitly evidenced by the infeasibility of various deterministic-mixture distributions in Sec. IV here. Entirely new techniques, therefore, are required to derive quantum-useful causal compatibility criteria for more general causal scenarios [6–8].

To this end, we introduce a technique applicable to general causal structures for deriving polynomial inequalities constraining observable probability distributions.

II. NOTATION AND DEFINITIONS

We follow the convention that upper-case letters indicate random variables while lower-case letters indicate some particular value associated with the corresponding random variable. In this convention, for example, a student's score on some exam X might depend probabilistically on the extent of sleep S. The Boolean proposition, or **event**, $X{=}x|S{=}s$ should be understood as "the students scores x on the exam given a duration of sleep equal to s. Events may be written in lower-case-only shorthand, such as $x|s$ instead of $X{=}x|S{=}s$.

Similarly, we indicate probability distributions using upper-case P, whereas lower-case p is used to indicate the probability of events. Thus $P(XY)$ is the multivariate probability distribution over the random variables $\{XY\}$, and $p(xy)$ is the joint probability of the two events $X{=}x$ *and* $Y{=}y$.

We choose to indicate logical negation by $\overline{x} \coloneqq \mathsf{Not}[x]$, such that $\overline{x}$ references the possibility of *any* outcome other than x, i.e. $p(\overline{x}) \coloneqq p(X{\neq}x) = 1 - p(X{=}x)$. This convention effectively coarse-grains the outcome space into binary possibilities.

A causal structure, for the purpose of this article, is taken to be *defined* by a directed acyclic graph (DAG): each node in the DAG corresponds to a random variable, each edge represents a causal influence between variables. In our graphical depictions we follow the convention of representing latent variables by circles, and observable variables by triangles [7].

The parents of a node X in a given graph G are defined as those nodes which have have directed edges originating at them and terminating at X, i.e. $\mathsf{Pa}_G(X)$ are all nodes $\boldsymbol{Y}$ such that for every $Y \in \boldsymbol{Y}$ there exists an edge from Y to X in G. Similarly the children of a node X in a given graph G are defined as those nodes which have have directed edges originating at X and terminating at them, i.e. $\mathsf{Ch}_G(X)$ are all nodes $\boldsymbol{Y}$ such that for every $Y \in \boldsymbol{Y}$ there exists an edge from X to Y in G. The **ancestors** of a node X in a given graph G are defined as those nodes which have a directed *path* to X, *including the node* X *itself*. Equivalently, the ancestors of X may be algorithmically accumulated by starting from X and then recursively enumerating all parents. $\mathsf{An}_G(\boldsymbol{V})$ means the ancestors of a *set* of random variables $\boldsymbol{V}$, $\mathsf{An}(\boldsymbol{V})$ is the union of all $\mathsf{An}(V) : V \in \boldsymbol{V}$, and so too for parents, children, etc.

$\mathsf{SubDAG}_G[\boldsymbol{V}]$ means the subgraph of the graph G corresponding to the nodes $\boldsymbol{V}$. $\mathsf{SubDAG}[\boldsymbol{V}]$ consists of the nodes $\boldsymbol{V}$ and the edges between pairs of nodes in $\boldsymbol{V}$ per the original graph. Of special importance to us is the **ancestral subgraph**, $\mathsf{AnSubDAG}[\boldsymbol{V}] \coloneqq \mathsf{SubDAG}[\mathsf{An}(\boldsymbol{V})]$, which is the minimal subgraph containing the full ancestry of $\boldsymbol{V}$.

A **node map** $f_{G,H}$ is a function which maps nodes in graph G to nodes in graph H. A node map should be thought of as a collection of rules about (some) of $\mathsf{Nodes}[G]$. A node map is generically *not* a graph map, i.e. $f(G) \neq H$.

A **complete node map** $f_{G,H}$ maps *every* $X \in \mathsf{Nodes}[G]$ to some node in $\mathsf{Nodes}[H]$. While the domain of f is the set of $\mathsf{Nodes}[G]$ whenever f is a complete node map, nevertheless the codomain of f need only be a subset of $\mathsf{Nodes}[H]$.

A **bijective node map** maps unique nodes of G to unique nodes of H. A bijective node map does not need to be complete; it may only explicitly dictate how a subset of $\mathsf{Nodes}[G]$ are mapped respective to a subset of $\mathsf{Nodes}[H]$.

A **graph isomorphism** $\mathrm{Iso}_{G,H}$ is both a complete bijective node map and a graph map which transforms G to H. Both properties are requires to be a graph isomorphism. That it to say, not only $\mathrm{Iso}(G) = H$ but also every $X \in \mathsf{Nodes}[G]$ is mapped to some unique $Y \in \mathsf{Nodes}[H]$ in one-to-one correspondence.

Node maps can be nested hierarchically: If the rules of map $f2$ are a subset of the rules of $f1$ then we say that $f2$ **respects** $f1$.

We now introduce the notion of DAG **inflation**. Consider an "original" DAG G, and an inflation of G which we will call $G' \in \mathsf{Inflations}[G]$. In order for G' to be a a "valid" inflation of G it must satisfy two properties:

1. There exists a **deflation map** *dmap*, i.e. a complete node map from the inflation DAG to the original DAG. *dmap* takes every node $X \in \mathsf{Nodes}[G']$ to some node in $\mathsf{Nodes}[G]$. To make this clear we demand that all those nodes $\boldsymbol{A} \in \mathsf{Nodes}[G']$ which map to A in G should *share the name* A, and be distinguished only by dummy indices A_1, A_2.... Using this labelling, we find that $\forall_i \; dmap(A_i){=}A$. The deflation map is surjective, and generally not bijective. We refer to the various A_i as multiple **copies** of A in the inflation DAG.
2. The deflation map must be a graph isomorphism at least when considering ancestral subgraphs within the inflation DAG. Formally: For all $X \in \mathsf{Nodes}[G']$ it must be true that $dmap(\mathsf{Pa}_{G'}(X)) = \mathsf{Pa}_G(dmap(X))$, i.e. the parents of every individual node in the inflation DAG G' map bijectively to the corresponding parents of the mapped node in the original DAG G[1].

The latter property is equivalent to $dmap(\mathsf{AnSubDAG}_{G'}[X]) = \mathsf{AnSubDAG}_G[dmap(X)]$ for all $X \in \mathsf{Nodes}[G']$. In other words, each individual random variable in the inflated DAG has precisely the same causal history as it *would have had* in the original DAG, up to the dummy indices indicating the deflation map.

The essential feature of an inflation DAG is that all the copies of a random variable are presumptively described by the same probability distribution: all the copies of an original random variable are identically distributed, by construction! This follows from assuming that copies of endogenous nodes in the inflated DAG are identically *and independently* distributed (i.i.d.), and that each copy of an exogenous node has the same functional dependence on its parents.

[Tobias' definition of a no-broadcasting inflation etc. has been moved to Sec. VII.]

III. THE INFLATION DAG TECHNIQUE FOR CAUSAL INFERENCE

The basic idea of our method is disarmingly simple, so let us illustrate it first with an example before presenting the general case in more formal detail. Consider the W-type distribution on the Triangle scenario,

$$p(001) = p(010) = p(100) = \frac{1}{3}.$$

So our question is, can this distribution over A, B and C be obtained through the causal structure of Fig. 1? For the sake of the argument, assume that this is the case; hence there are distributions over the hidden variables $p(X)$, $p(Y)$ and $p(Z)$ as well as conditional distributions $p(A|X,Y)$, $p(B|Y,Z)$ and $p(C|X,Z)$ such that $p(ABC)$ is reproduced. Intuitively, we may think of X as a "black box" system that outputs random stuff, and likewise of A as a black box which takes some inputs and outputs random signals. Now there are two things that we can do in order to

- Of each outgoing variable we can produce an arbitrary number of further copies, resulting in new edges;
- Of each black box, we can likewise produce copies that behave independently, resulting in new nodes.

We can then rewire these gadgets into a new Bayesian network. Fig. 2 presents an example of such a rewiring, or "inflated" DAG. Pearl's *do calculus* uses rewirings of a similar nature, and also Henson, Lal and Pusey [7] use this reasoning to show that the entropic inequality $I(A:B) + I(B:C) \leq H(B)$ holds in all GPTs, just without taking copies of variables but only rewiring the boxes.

Since the outcome $B = 1$ occurs sometimes, there must be hidden variables values a_2 and c_3 such that $B_4 = 1$ has positive probability; and similarly, we choose values for X_1 and Y_1 such that $A_1 = 1$ has positive probability, and for X_2 and Z_1 such that $C_2 = 1$ has positive probability. Since all the hidden variables in the inflated DAG are independent, these conditions do sometimes all occur jointly. Then by the form of the assumed distribution, it follows that for the assumed values of the hidden variables, one must necessarily have $A_2 = 0$, $B_3 = 0$ and $C_1 = 0$, which contradicts the assumption $p(000) = 0$.

---

[1] This is reminiscent of covering spaces in topology.

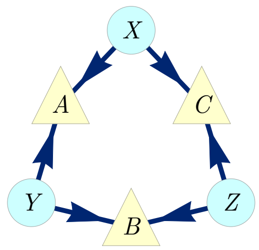

FIG. 1. The causal structure of the Triangle scenario.

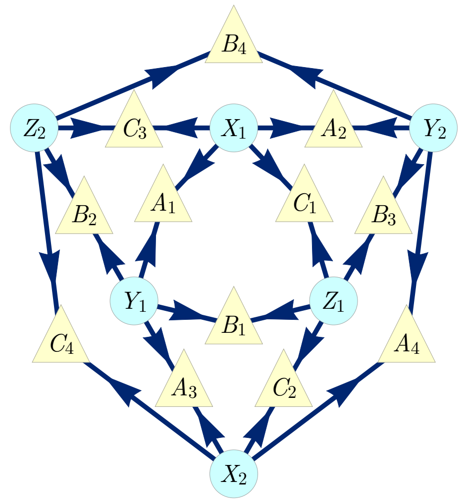

FIG. 2. An inflation DAG of the Triangle scenario, where each root node has been duplicated.

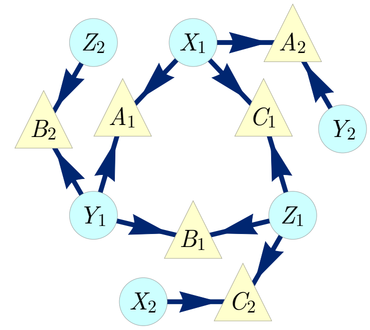

FIG. 3. A inflation DAG of Triangle scenario consisting of 6 inflated observables and their complete causal histories, which is a subgraph of Fig. 2.

It is not difficult to transform this argument into a polynomial inequality for the Triangle scenario. However, it is more instructive to present the derivation of polynomial inequalities using our method in a slightly different manner, and we will do so later.

Consider the Triangle scenario [7 (Fig. E#8), 5 (Fig. 18b), 6 (Fig. 3), 10 (Fig. 6a), 17 (Fig. 1a), 18 (Fig. 8), 19 (Fig. 1b), 11 (Fig. 4b)], the causal structure of which is depicted here in Fig. 1. The Triangle scenario is a correlation scenario in the sense of Fritz [6]; see especially Sec. 2.3 there.

Possible inflations of the Triangle scenario are depicted in Figs. 2 and 3; every copy of the random variable A in an inflation has precisely two parents, one of which is a copy of the random variable X and one of which is a copy of the random variable Y, just as in the original DAG. The duplication of Y into Y_1 and Y_2 leads to X_1 having outgoing edges to both A_1 and A_2 in Figs. 2 and 3. Thus, inflating a DAG not only leads to duplicated nodes but to duplicated *edges*.

Duplicating an edge in a causal structure means broadcasting information about the state of a random variable; in our example, the information about X which was "sent" to A is effectively broadcast to both A_1 and A_2 in the inflation. This is quite intentional. Quantum theory is governed by a no-broadcasting theorem [20, 21]; by electing to embed broadcasting into an inflation DAG we can specifically construct a foil to quantum causal structures. This contrast is elaborated further in Sec. VII.

By imagining an inflated DAG we can consider probabilities for new kinds of multivariable events, such as $p(A_1{=}0 \wedge A_2{=}1)$. It is convenient to have a shorthand notation for events which refer to multiple copies of a random variable. We therefore introduce special notation based on subscript-indexing of the lower-case value indicators.

$$\text{Let}\quad p(a_1 a_2 ...) \coloneqq p(A_1{=}a \bigwedge A_2{=}a \bigwedge ...) \quad\text{and}\quad p(\overline{a_1} a_2 ...) \coloneqq p(A_1{\neq}a \bigwedge A_2{=}a \bigwedge ...) \quad \text{etc.}$$

To be perfectly clear however, a_1 and a_2 *do not* refer to two distinct possible outcomes of one random variable; rather they represent the event in which two copies of a random variable *both* have the same outcome a. Moreover, generally $p(a_1 a_2) \neq p(a)$, because although A_1 and A_2 are identically distributed they may not be perfectly correlated. In the same vein, generally $p(a_1 a_2) \neq p(a_1)p(a_2)$, because although A_1 and A_2 are identically distributed they may not be independently distributed. Indeed, in Fig. 3, for example, A_1 and A_2 share the common ancestor X_1, and hence they are not independent. On the other hand, sometimes non-root copies of a random variable *can* be i.i.d., such as A_1 and A_4 in Fig. 2.

The discussion which follows uses this new notation, and relies extensively on this definition of an inflated DAG. This enables us to explain the four-step process one can use in order to derive polynomial infeasibility criteria pertaining to original DAG from properties of the inflated DAG.

Step 1: **Generate an initial set of linear inequalities**

In order to derive inequalities on the original DAG, we begin by collecting a generating set of linear conditions pertaining to the observable variables in the inflated DAG. The generating set is identically nonnegativity of all non-latent probabilities. For simplicity we take all observable variables to be binary, but the derivation can easily be adjusted to account for any number of outcomes. Taking the six observable variables in Fig. 3 to be binary, however, lead to 64 distinct nonnegativity conditions, corresponding to $0 \leq P(A_1A_2B_1B_2C_1C_2)$. The starting nonnegativity inequalities therefore look like

$$
\begin{aligned}
&0 \leq p(a_1a_2b_1b_2c_1c_2)\\
&0 \leq [p(\overline{a_1}a_2b_1b_2c_1c_2) =]\, p(a_2b_1b_2c_1c_2) - p(a_1a_2b_1b_2c_1c_2)\\
&0 \leq [p(a_1\overline{a_2}b_1b_2c_1c_2) =]\, p(a_1b_1b_2c_1c_2) - p(a_1a_2b_1b_2c_1c_2)\\
&\vdots\\
&0 \leq [p(\overline{a_1a_2}b_1b_2c_1c_2) =]\, p(b_1b_2c_1c_2) - p(a_1b_1b_2c_1c_2) - p(a_2b_1b_2c_1c_2) + p(a_1a_2b_1b_2c_1c_2)\\
&0 \leq \left[p(\overline{a_1}a_2\overline{b_1}b_2c_1c_2) =\right] p(a_2b_2c_1c_2) - p(a_1a_2b_2c_1c_2) - p(a_2b_1b_2c_1c_2) + p(a_1a_2b_1b_2c_1c_2)\\
&\vdots
\end{aligned}
\tag{1}
$$

etc. These inequalities are found by iterating over the general definition of $p(\boldsymbol{x}\overline{\boldsymbol{y}}) = p(\boldsymbol{x}) - p(\boldsymbol{x}\boldsymbol{y})$ applied to all possible joint probabilities.

Step 2: **Identify factorization of probabilities from the inflated DAG**

These "trivial" linear inequalities on the inflated DAG can be made into (weak) nontrivial polynomial inequalities by accounting for the marginal independence of certain random variable subsets. Independence of various distributions implies factorization of joint probabilities; the factorization of probabilities is the first step in transforming our set of otherwise-linear conditions.

We inspect the inflated DAG to find pairwise **causal independence relations** in order to identify all subset marginal independences. We define a pair of random variables U and V to be causally independent if and only if they do not share any common ancestor, i.e. $U \hookleftarrow\text{no-comm-an-w}/\hookrightarrow V$. No-common-ancestor is equivalent to d-separation by the empty set [1–4]. Since we we consider all feasible distributions to be globally Markov with respect to the DAG, it follows that causally independent variables must have marginally independent distributions. In other words, no statistically correlation without causal explanation.

The six pairs of causally-independent variables in Fig. 3, and some of their marginal independence implications, are given by

$$
\begin{array}{l}
C_2 \hookleftarrow\text{no-comm-an-w}/\hookrightarrow A_1 \text{ or } A_2 \text{ or } B_2\\
A_2 \hookleftarrow\text{no-comm-an-w}/\hookrightarrow B_1 \text{ or } B_2\\
B_2 \hookleftarrow\text{no-comm-an-w}/\hookrightarrow C_1
\end{array}
\quad \text{implies that} \quad
\begin{array}{l}
P(A_1A_2B_2C_2) = P(C_2) \times P(A_1A_2B_2)\\
P(A_2B_1B_2C_2) = P(A_2) \times P(B_1B_2C_2)\\
P(A_2B_2C_1C_2) = P(B_2) \times P(A_2C_1C_2)\\
P(A_2B_2C_2) = P(A_2) \times P(B_2) \times P(C_2)
\end{array}
\tag{2}
$$

etc. Interestingly, there are *zero* observable unconditional (or even conditional!) independence relations implied by the original DAG in Fig. 1.

By accounting for marginal independence relations, the set of linear inequalities is transformed into a set of polynomial inequalities. Eqs. (2) implies the factorization of some many different probabilities, such as $p(a_2b_1b_2c_2) \to p(a_2)p(b_1b_2c_2)$, $p(a_2b_1b_2) \to p(a_2)p(b_1b_2)$, $p(a_2b_1) \to p(a_2)p(b_1)$, etc.

Step 3: **Map marginal distributions in the inflated DAG to those in the original structure**

Recall that the random variables in the inflated DAG are constructed to be copies of the original random variables. As such, it is possible that two distinct subgraphs of the inflated DAG may represent precisely the *same* causal structure, if the two subgraphs are equivalent up to "dummy index" of the copied variables. Moreover, some subgraphs of the inflated DAG are equivalent to subgraphs of the original DAG, in the same sense of equivalent. When two graphs are equivalent in this sense we call them **copy equivalent**. Note that this notion of copy equivalence is stronger than merely graph isomorphism: For two causal structures to be copy equivalent there must exist an isomorphism between the graphs such that every random variables is mapped to itself or a copy of itself. With expanded notation moved up front, this discussion should be trimmed.

Consider, now, two different subsets of observable random variables in the inflated DAG. Both subsets should be described by the same marginal probability distribution if the **ancestral subgraphs** of both sets are copy equivalent. We define the ancestral subgraph of a set of variables $\boldsymbol{V}$ to mean the subgraph of the causal structure containing $\boldsymbol{V}$'s complete causal history, i.e. $\mathsf{Ancestors}\big[\boldsymbol{V}\big]$. $\mathsf{Ancestors}\big[\boldsymbol{V}\big]$ refers to the set $\boldsymbol{V}$ along with any vertices that have a directed path to leading from them to $\boldsymbol{V}$. Thus the ancestral subgraph of $\boldsymbol{V}$ consists of the vertices $\mathsf{Ancestors}\big[\boldsymbol{V}\big]$ and all edges between them per the original graph. As an example: Fig. 3 is the ancestral subgraph of Fig. 2 on the six observable variables $\{A_1A_2B_1B_2C_1C_2\}$. The ancestral subgraph corresponding to observables $\{A_1B_2C_1\}$ is illustrated in Fig. 4. In contrast, the ancestral subgraph of $\{A_1B_1C_1\}$ is copy equivalent to Fig. 1. When a set of variables admits partitioning into causally independent partitions, then the ancestral subgraph of that set will have unconnected components. An example of a multiple-components ancestral subgraph is Fig. 5, where $A_2 \perp \{B_1C_2\}$, as per Eqs. (2).

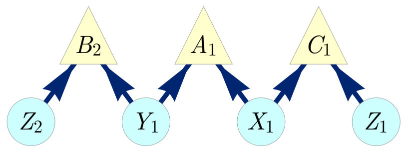

FIG. 4. A single-component ancestral subgraph of both Figs. 2 and 3, consisting of observables $\{A_1B_2C_1\}$ and their complete causal histories.

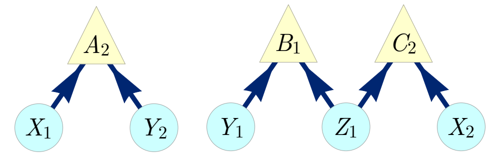

FIG. 5. A multiple-components ancestral subgraph of both Figs. 2 and 3, consisting of observables $\{A_2B_1C_2\}$ and their complete causal histories.

In sum: If two sets of random variables, $\boldsymbol{S}$ and $\boldsymbol{T}$, are equivalent up to their dummy indices AND the ancestral subgraphs of $\boldsymbol{S}$ and $\boldsymbol{T}$ are copy equivalent, then $P(\boldsymbol{S}) = P(\boldsymbol{T})$. By inspecting different ancestral subgraphs for copy equivalence we can establish perfect equivalence of various marginal distributions. Coinciding marginal distributions reduces the number of parameters needed to describe the joint distribution over the entire inflated DAG. Multiple superficially distinct probabilities in the inflated DAG may actually share a single parameter. As an example, note that the marginal distributions over $\{A_1A_2B_1\}$ and $\{A_1A_2B_2\}$ must coincide, due to their copy-equivalent ancestral subgraphs. Consequently $p(a_1a_2b_1) = p(a_1a_2b_2)$, $p(\overline{a_1}a_2b_1) = p(\overline{a_1}a_2b_2)$, etc. The number of equivalent *probabilities* implied by one pair of coinciding *distributions* pertains to the (multivariate) range of the distributions. All the coincidences of marginal distributions implied by Fig. 3 are given by

$$P(A_1A_2B_1) = P(A_1A_2B_2),\ P(B_1B_2C_1) = P(B_1B_2C_2),\ P(A_1C_1C_2) = P(A_2C_1C_2), \tag{3}$$

and their subsets.

To determine all marginal distribution coincidences we constructed the ancestral subgraph of every subset of observable random variables in the inflated DAG, and then picked out all copy equivalent pairings[2]. We note that Step 2 of finding marginally independent subsets can also be recast in terms of ancestral subgraphs: Sets of random variables $\boldsymbol{S}$ and $\boldsymbol{T}$ must be marginally independent if the ancestral subgraph of $\boldsymbol{S}$ does not intersect the ancestral subgraph of $\boldsymbol{T}$. This is illustrated in Fig. 5, where the ancestral subgraph of A_2 and the ancestral subgraph of $\{B_1C_2\}$ comprise disconnected components in the ancestral subgraph of $\{A_2B_1C_2\}$.

In addition to such internal equivalency relations, we can further identify some marginal multivariate distributions in the inflated DAG which correspond to marginal distributions in the original DAG. The set of *inter-DAG* marginal distribution coincidences may provide partial information regarding the set of *intra-DAG* marginal distribution coincidences; i.e. the set of inter-DAG probabilities-mapping relations effectually captures a subset of the intra-DAG probabilities-equivalency relations. The inter-DAG marginal distribution coincidences between Figs. 1 and 3 are

$$P(A_1B_2) = P(AB),\ P(B_1C_2) = P(BC),\ P(A_2C_1) = P(AC),\ P(A_1B_1C_1) = P(ABC), \tag{4}$$

and their subsets.

The inter-DAG probabilities-mapping relations allow us to transform our set of inequalities once more, this time into inequalities which have bearing on the original causal structure. The inter-DAG probabilities-mapping relations and intra-DAG probabilities-equivalency relations implied by Fig. 3, per Eqs. (3,4), are

$$\underbrace{\begin{array}{l}\{p(a_1)=p(a_2)\}\to p(a)\\ \{p(b_1)=p(b_2)\}\to p(b)\\ \{p(c_1)=p(c_2)\}\to p(c)\end{array}}_{\text{Map per \textit{definition} of inflation DAG}}\quad \underbrace{\begin{array}{l}\{p(a_1b_1)=p(a_1b_2)\}\to p(ab)\\ \{p(a_1c_1)=p(a_2c_1)\}\to p(ac)\\ \{p(b_1c_1)=p(b_1c_2)\}\to p(bc)\\ p(a_1b_1c_1)\to p(abc)\end{array}}_{\text{Map from inflated DAG to original}}\quad \underbrace{\begin{array}{l}\{p(a_1a_2b_1)=p(a_1a_2b_2)\}\to \text{n/a}\\ \{p(a_1c_1c_2)=p(a_2c_1c_2)\}\to \text{n/a}\\ \{p(b_1b_2c_1)=p(b_1b_2c_2)\}\to \text{n/a}\end{array}}_{\text{Other intra-DAG equivalency relations}} \tag{5}$$

After applying the map we are left we a system of **hybrid inequalities** of sorts, simultaneously containing two radically different kinds of probabilities. Some probabilities now pertain to the original DAG, but they appears alongside many probabilities which *do not map* to the original DAG. Probabilities which cannot be related to the original causal structure include $\{p(a_1b_1c_2), p(a_1b_2c_1), p(a_2b_1c_1)\}$, and any joint probability which references more than one instance of a duplicated variable, such as $\{p(a_1a_2), p(a_1a_2b_1), ...\}$.

We refer to those joint probabilities pertaining to the inflated DAG which have no parallel in the original DAG as **gedankenprobabilities**, as they could be measured in principle if one were to construct the inflated causal structure.

---

[2] Actually, the authors used a somewhat more efficient method to infer all marginal distribution coincidences. Discussing optimization is unwarranted, though, because this task is so easily accomplished by a computer.

As we are really only concerned with the original DAG, however, these "unmeasurable" joint distributions are effectively just thought experiments. The in-principle existence of gedankenprobabilities, however, is critical to inferring causal infeasibility criteria for the original DAG. Interestingly, gedankenprobabilities only exist in the classical probability theory; their non-existence in the context of quantum causal structures is discussed in Sec. VII.

The set of observable multivariate distributions in the inflated DAG of Fig. 3 which *do not* coincide with any observable distribution in the original DAG of Fig. 1 consists of

$$P(A_1A_2),\ P(B_1B_2),\ P(C_1C_2),\ P(A_1B_1C_2),\ P(A_1B_2C_1),\ P(A_2B_1C_1) \tag{6}$$

and superset-distributions of those in Eq. (6), such as $P(A_1A_2B_1)$ etc.

The fact that $p(a_1b_2c_1) \neq p(abc)$ is an important subtlety, and is a consequence of the fact that the ancestral subgraphs of $\{A_1B_2C_1\}$ and $\{ABC\}$ are not copy equivalent; indeed, their ancestral subgraphs are not even isomorphic.

Step 4: **Quantifier elimination of the gedankenprobabilities**

We can infer implications for the original random variable from the system of hybrid inequalities obtained after Step 3. This inference task is essential a form of quantifier elimination. Just the *existence* of gedankenprobabilities such as $p(a_1a_2)$ can be used to imply constraints on the relevant probabilities $p(a)$, $p(b)$, $p(c)$, $p(ab)$, $p(ac)$, $p(bc)$, $p(abc)$. Thus, the final step toward obtaining the desired causal infeasibility criteria is to eliminate the gedankenprobabilities from our system of hybrid polynomial inequalities. This quantifier elimination problem is well defined mathematically, although it is a challenging problem when the quantifiers are nonlinear.

Many modern computer algebra systems have functions capable of tackling this sort of problem fully symbolically[3]. One might then hope to use such software systems to rid the hybrid inequalities of the gedankenprobabilities. Currently, however, it is not practical to perform nonlinear quantifier elimination on large polynomial systems with many quantifiers. We consider, therefore, two other strategies for making effective use of the hybrid inequalities.

Firstly, one may substitute numeric values for all the joint probabilities pertaining to the original DAG which appear in the polynomial inequality set. Upon doing so, the quantifier elimination problem is converted to a quantifier existence problem: Do there exist gedankenprobabilities that satisfy the otherwise-numeric system of polynomial inequalities? Most computer algebra systems can resolve such yes/no *satisfiability* questions quite rapidly[4].

Note that real-world data with uncertainties can also be incorporated into these satisfiability questions. Instead of asserting that a particular probability is equal to a *number*, one can incorporate new inequalities which constrain the experimentally-known probabilities to lie in given *intervals*. Assigning probabilities to intervals as opposed to numeric values results in further free parameters in the system, but the problem nevertheless remains one of *universal* existential closure, and can be efficiently tested.

The second strategy is to relax the polynomial quantifier elimination problem to a linear one. This is accomplished by treating products of variables as independent variables. When encoding the polynomial inequalities into a linear solver, so as to use linear quantifier elimination methods, we must specify the variables which are being eliminated. From the nonlinear perspective, the system of inequalities resulting from Fig. 3 contains 34 qedankenprobabilities per Eq. (6); from the linear perspective, however, there are 6 additional "terms" to be eliminated from the system.

$$\underbrace{p(a_1a_2),\ldots,\ p(a_1b_2c_1),\ldots,\ p(a_1a_2b_1b_2c_1c_2)}_{\text{Gedankensprobabilities, quantifiers for nonlinear elimination}} \qquad \underbrace{p(a)p(b_1b_2),\ldots,\ p(c)p(a_1a_2),\ldots,\ p(c)p(a_1a_2b_2)}_{\text{Further quantifiers vis–à–vis linear elimination}}$$

This strategy allows us to obtain polynomial inequalities in a computationally efficient fashion. Ref. [23] similarly treats products as independent quantifiers for linear elimination. To be clear, if we *linearly* eliminate $p(a_1a_2)$ from our system of inequalities we would *still* have terms such as $p(a_1a_2)p(c_2)$ and $p(a_1a_2b_2)p(c_2)$.

Geometrically, linear quantifier elimination is equivalent to projecting a high-dimensional polytope described by halfspaces (inequalities) into a lower-dimensional subspace. Polytope projecting is a well-understood problem in computational optimization, and a surprising variety of algorithms are available for the task [24–26]. The oldest-known method for polytope projection, i.e. linear quantifier elimination, is an algorithm known as Fourier-Motzkin (FM) elimination, although Block Elimination and Vertex Enumeration [27] are also fairly popular. Much more advanced polytope projection algorithms, such as Equality Set Projection (ESP) and Parametric Linear Programming, have also recently become available [24–26].

---

[3] For example *Mathematica*^{TM}'s `Resolve` command, *Redlog*'s `rlposqe`, or *Maple*^{TM}'s `RepresentingQuantifierFreeFormula`, etc.

[4] The undocumented *Mathematica*^{TM} function `` Reduce`ExistsRealQ `` is quite effective for such purposes, and specialized satisfiability software such as SMT-LIB's `check-sat` [22] are also well-suited for this purpose. Also, any numeric optimizer which accepts nonlinear constraints can also be used for nonlinear quantifier existence testing: the optimizer will return an error message when a set of constraints cannot be satisfied. Nonlinear optimizers include *Maple*^{TM}'s `NLPSolve`, *Mathematica*^{TM}'s `NMinimize`, and dozens of free and commercial optimizers for `AMPL` and/or `GAMS`

Linear quantifier elimination routines are available in many software tools[5]. The authors elected to custom-code an FM elimination routine in *Mathematica*^{TM}[6]. Fourier-Motzkin elimination is the worst possible polytope projection algorithm to use, however, when the dimension of the final projection is much smaller than the initial space of the inequalities, i.e. when there are many gedankenprobabilities. See Refs. [24–26] and Appendix A for further detail.

Quantifier elimination is already widely used in causal inference to derive entropic inequalities [9–11]. In that task, however, the quantifiers being eliminated are those entropies which refer to hidden variables. By contrast, the probabilities we consider here are exclusively in terms of observable variables right from the very start, the inequalities are always latent-variable free. The quantifiers we eliminate are the gedankenprobabilities, which are quite different from probabilities involving hidden variables.

Although linear quantifier elimination can be highly optimized, it can still prove computationally difficult. One useful alternative to linear quantifier elimination is to identify representative probability distributions which are incompatible with the hybrid inequalities; Chaves and Budroni [28] use this technique, for example.

A final alternative to linear quantifier elimination is to restrict one's consideration to quantifier-free inequalities with a particular form. We found this alternative technique — trading generality for speed — to be extraordinarily practical. The subtype of causal criteria which which can be most rapidly recognized are those which follow from certain tautologies in classical propositional logic, see Sec. VIII for further details.

TABLE I. A comparison of different approaches to utilizing nonlinear inequalities containing gedankenprobabilities. The primary divide is quantifier elimination versus quantifier existence, with approaches being further subdivided into linear and nonlinear variants. The alternate approach based on a logical tautologies, discussed in Sec. VIII, is technically not a utilization of hybrid inequalities. It is included here only for comparison, as it is also a way of obtaining quantifier-free polynomial inequalities.

Approach	Also Known As	Difficulty	End Results
Nonlinear Quantifier Elimination	Resolving Partial Existential Closure	Very Hard	Necessary-but-not-Sufficient Polynomial Inequalities
Nonlinear Quantifier Satisfiability	Universal Existential Closure, Nonlinear Programming	Easy	Certify the infeasibility of a specific distribution
Linear Quantifier Elimination	Polytope Projection	Moderate	Necessary-but-not-Sufficient Polynomial Inequalities
Linear Quantifier Satisfiability	Linear Programming, Universal Existential Closure	Very Easy	Certify the infeasibility of a specific distribution
Alternate: Logical Tautologies	Combinatorial / Set-Theoretic Implications	Easy	Necessary-but-not-Sufficient Polynomial Inequalities

It is also possible to use a mixed strategy of linear and nonlinear quantifier elimination, such as Chaves [23] advocates. It is interesting to compare the technique for deriving polynomial inequalities here to that in in Ref. [23]. One the one hand, our initial set of linear inequalities is much stronger, as we work with the inflated DAG whereas Chaves [23] considers only the original DAG. This allows us to analyze scenarios which Ref. [23] cannot, namely those without any observable conditional independence relations, such as the Triangle scenario. On the other hand, Ref. [23] purportedly incorporates any kind of observable conditional independence relation, whereas we account only for *unconditional* independence relations, per Step 2. A careful examination of Ref. [23], however, reveals that only unconditional independence relations are utilized in all the example there. To that extent, therefore, all the explicit results in Ref. [23] are implied by the inflation DAG technique.

As far as we can tell, our inequalities are not related to the nonlinear causal infeasibility which have been derived specifically to constrain classical networks [29–31], nor to the nonlinear inequalities which account for interventions to a given causal structure [32, 33].

IV. EXAMPLES OF POLYNOMIAL INEQUALITIES AND INFEASIBLE DISTRIBUTIONS WHICH VIOLATE THEM

Here are some examples of causal infeasibility criteria for the Triangle scenario which we can derive by considering inflation DAGs.

---

[5] For example *MATLAB*^{TM}'s `MPT2`/`MPT3`, *Maxima*'s `fourier_elim`, *lrs*'s `fourier`, or *Maple*^{TM}'s (v17+) `LinearSolve` and `Projection`. Frustratingly, almost all these software tools currently rely on plain FM elimination as the underlying algorithm. The computational complexity of FM scales with the number of quantifiers being eliminated [24, 25]. `MPT2` did support ESP, but had restrictive limits on the dimension of the projection. The python module *polytope* (v0.1.1 2015-10-26) has the (undocumented) option of projection via ESP.

[6] FM elimination algorithms make intermittent calls to a linear-programming subroutine for eliminating redundant inequalities. The authors found an efficient implementation of this subroutine in *Mathematica*^{TM}, see Appendix B for further details.

The nontrivial polynomial inequality

$$p(a) + p(b) + p(c) \leq 1 + p(ab) + p(ac) + p(b)p(c) \tag{7}$$

is found to be a constraint on the Triangle scenario through by summing the follow two inequalities

$$\begin{aligned} &0 \leq 1 - p(a_1) - p(b_2) - p(c_1) + p(a_1b_2) + p(a_1c_1) + p(b_2c_1) - p(a_1b_2c_1) \quad \left[= p(\overline{a_1}\overline{b_2}\overline{c_1})\right] \\ &0 \leq p(a_1b_2c_1) \end{aligned} \tag{8}$$

subject to the following transformations

$$\underbrace{p(b_2c_1) \to p(b_2)p(c_1)}_{\text{Factorization relation, per Eqs. (2).}} \qquad \text{and} \qquad \underbrace{\begin{array}{c} p(a_1) \to p(a)\,,\ p(b_2) \to p(b)\,,\ p(c_1) \to p(c), \\ p(a_1b_2) \to p(ab),\ p(a_1c_1) \to p(ac) \end{array}}_{\text{Mapping relations, per Eqs. (5)}}. \tag{9}$$

Those transformation follow from fundamental properties the inflation DAG's marginal distributions, namely

$$P(B_2C_1) = P(B_2) \times P(C_1), \quad P(A_1B_2) = P(AB), \quad P(A_1C_1) = P(AC). \tag{10}$$

Indeed, this example has been chosen because the properties in Eq. (10) can be easily be visually verified from the small ancestral subgraph of $\{A_1B_2C_1\}$, namely Fig. 4.

A consequence of Eq. (7) is that the GHZ-type distribution

$$P_{\text{GHZ}}(ABC) \coloneqq \quad p_{\text{GHZ}}(abc) = \frac{[000] + [111]}{2} = \begin{cases} \frac{1}{2} & \text{if } \ a = b = c \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

is found to be infeasible from the Triangle scenario, by considering the special case of Eq. (7) where $a{\to}0$, $b{\to}1$, $c{\to}1$. Earlier works have already shown that the GHZ-type distribution is incompatible with the classical Triangle scenario [6, 10, 19]. Interestingly, however, the entropic monogamy relation $I(A:B) + I(A:C) \leq H(A)$ which rejects the GHZ-type distribution has been shown also to hold if the hidden shared resources are non-classical, even using generalized probabilistic theories [7, Cor. 24].

Slightly more involved but otherwise analogous considerations give rise to the (pre-transformations) inequality

$$p(a_2b_2c_2) \leq p(\overline{a_1}\overline{b_1}\overline{c_1}) + p(a_1b_2c_2) + p(a_2b_1c_2) + p(a_2b_2c_1) \tag{12}$$

which in turn yields

$$p(a)p(b)p(c) \leq p(\overline{a}\overline{b}\overline{c}) + p(c)p(ab) + p(b)p(ac) + p(a)p(bc) \tag{13}$$

per

$$\underbrace{\begin{array}{l} p(a_1b_2c_2) \to p(c_2)p(a_1b_2) \\ p(a_2b_1c_2) \to p(a_2)p(b_1c_2) \\ p(a_2b_2c_1) \to p(b_2)p(a_2c_1) \\ p(a_2b_2c_2) \to p(a_2)p(b_2)p(c_2) \end{array}}_{\text{Factorization relations}} \qquad \text{and} \qquad \underbrace{\begin{array}{l} p(a_1b_2) \to p(ab) \\ p(a_2c_1) \to p(ac) \\ p(b_1c_2) \to p(bc) \\ p(\overline{a_1}\overline{b_1}\overline{c_1}) \to p(\overline{a}\overline{b}\overline{c}) \end{array}}_{\text{Nontrivial mapping relations}}. \tag{14}$$

A consequence of Eq. (13) is that the W-type distribution

$$P_{\text{W}}(ABC) \coloneqq \quad p_{\text{W}}(abc) = \frac{[100] + [010] + [001]}{3} = \begin{cases} \frac{1}{3} & \text{if } \ a + b + c = 1 \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

is found to be infeasible from the Triangle scenario, by considering the special case of Eq. (13) where $a{\to}1$, $b{\to}1$, $c{\to}1$.

V. FAILURE OF ENTROPIC INEQUALITIES

It is interesting to note that entropic inequalities [9–11] fail to detect the infeasibility of the W-type distribution [Eq. (15)] with respect to the Triangle scenario [Fig. 1]. The polynomial inequality (13), by contrast, is a better causal infeasibility criterion in that it successfully rejects the W-type distribution.

The entropic inequalities associated with the Triangle scenario are given by

$$\begin{aligned} & I(A:B)+I(A:C) \leq H(A) \\ \text{and} \quad & I(A:B)+I(A:C)+I(B:C) \leq H(A,B)-I(A:B:C) \\ \text{and} \quad & I(A:B)+I(A:C)+I(B:C) \leq \frac{H(A)+H(B)+H(C)}{2}-I(A:B:C) \end{aligned} \tag{16}$$

and their permutations [7, 10, 17, 19].

Bipartite mutual information is defined as as $I(A:B) \coloneqq H(A)+H(B)-H(A,B)$ and tripartite mutual information is defined as $I(A,B,C) \coloneqq H(A)+H(B)+H(C)-H(A,B)-H(A,C)-H(B,C)+H(A,B,C)$. We find, then, that the W-type distribution satisfies all of the inequalities in Eq. (16).

Entropic inequalities are widely considered state-of-the-art causal infeasibility criteria [7]. They are well known to be necessary but not sufficient characterizations of causal structures. The insufficiency of entropic inequalities is a pressing concern in quantum information theory since they often fail to detect the infeasibility of quantum distributions, for instance in the Triangle scenario [6, Prob. 2.17].

Replacing the classical hidden variables in a causal structure with quantum states results in a quantum causal structure through which more distributions are feasible than would be classically. The quantum distributions are superficially quite similar to their classical subset. For example, both sets respect all conditional independence relations implied by the common underlying causal structure. Recent work has found that quantum causal structure imply many of the entropic inequalities implied by their classical counterparts [7, 17, 28]. Admittedly, some quantum correlation are identified by entropic inequalities [34, 35], but many are not [6]. The superficial similarity between quantum and classical distributions demands especially sensitive causal infeasibility criteria in order to distinguish them; our polynomial inequalities appear to be suitable tools for this purpose.

VI. THE BIPARTITE BELL SCENARIO

To further illustrate our inflation-DAG approach for deriving causal infeasibility criteria we demonstrate how to recover the original Bell inequalities [12, 15, 36] via our method. It is critical that our method should be able to derive these seminal criteria, as Bell inequalities have been a foundational component of quantum information theory for the last half century [14, 37].

Consider the causal structure associated to the Bell/CHSH [12, 15, 36, 38] experiment [7 (Fig. E#2), 5 (Fig. 19), 10 (Fig. 1), 8 (Fig. 1), 39 (Fig. 2b), 33 (Fig. 2)], depicted here in Fig. 6. $\{A, B, X, Y\}$ are all observable variables, and Λ is the latent common cause of A and B.

Analysis of the inflated DAG in Fig. 7 shows that

$$p(abxy)p(\bar{x})p(\bar{y}) \leq p(a\bar{b}x\bar{y})p(\bar{x})p(y)+p(\bar{a}b\bar{x}y)p(x)p(\bar{y})+p(ab\bar{x}\bar{y})p(x)p(y) \tag{17}$$

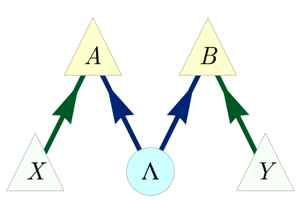

FIG. 6. The causal structure of the a bipartite Bell scenario. The local outcomes of Alice's and Bob's experimental probing is assumed to be a function of some latent common cause, in addition to their independent local experimental settings.

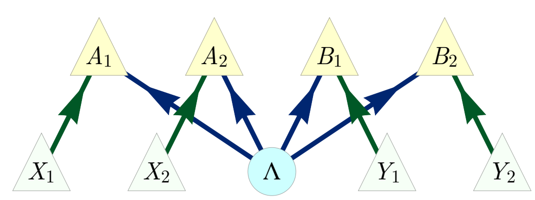

FIG. 7. An inflation DAG of the bipartite Bell scenario, where both local settings variables have been duplicated.

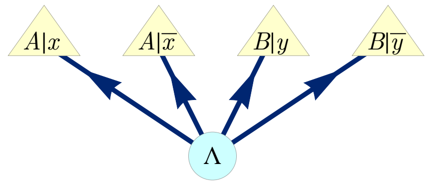

FIG. 8. The causal structure of the Bell scenario expressed in a form which makes use of conditional random variables.

by virtue of the unmapped (but factored) inequality

$$p(a_1 b_1 x_1 y_1) p(\overline{x_2}) p(\overline{y_2}) \leq p(a_1 \overline{b_2} x_1 \overline{y_2}) p(\overline{x_2}) p(y_1) + p(\overline{a_2} b_1 \overline{x_2} y_1) p(x_1) p(\overline{y_2}) + p(a_2 b_2 \overline{x_2} \overline{y_2}) p(x_1) p(y_1) \tag{18}$$

where *every* probability appearing in Eq. (18) maps to the original scenario, hence yielding Eq. (17). To derive the usual Bell inequalities from Eq. (17) we switch to conditional probabilities via $p(abxy) \to p(ab|xy)p(xy) = p(ab|xy)p(x)p(y)$ which, after dividing both sides of Eq. (17) by $p(x)p(\overline{x})p(y)p(\overline{y})$, yields

$$p(ab|xy) \leq p(a\bar{b}|x\overline{y}) + p(\overline{a}b|\overline{x}y) + p(ab|\overline{x}\overline{y}) \tag{19}$$

or, equivalently,

$$p(ab|xy) + p(ab|x\overline{y}) + p(ab|\overline{x}y) \leq p(a|\overline{x}) + p(b|\overline{y}) + p(ab|\overline{x}\overline{y}) \tag{20}$$

which is precisely the Clauser-Horne (CH) inequality [40] for the Bell scenario. Note that to obtain Eq. (20) from Eq. (19) we implicitly made of of the no-signalling assumptions, namely $p(a|xy) = p(a|x)$ and $p(b|xy) = p(b|y)$. The CH inequality is the *unique* Bell inequality (up to permutations) for the Bell scenario if $\{A, B, X, Y\}$ are all binary, and hence the CH inequality is a necessary and sufficient criterion to ascertain if correlations are compatible with that Bell scenario variant.

The causal structure of a Bell scenario can also be formulated directly in terms of conditional random variables. For example, the conditional-structure interpretation of Fig. 7 is Fig. 8.

The Bell inequalities are then self-evident from Fig. 8 without the need for an inflation DAG. The conditional-structure formulation innately implies its own inaccessible gedankenprobabilities, such as $\{p(a|x, a|\overline{x}),\, p(a|x, \overline{a}|\overline{x}), ...\}$ etc. By eliminating these gedankenprobabilities from the set of inequalities generated by $0 \leq P(A|x, A|\overline{x}, B|y, B|\overline{y})$ we obtain

$$0 \leq p(a|\overline{x}) + p(b|\overline{y}) + p(a|x, b|y) - p(a|x, b|\overline{y}) - p(a|\overline{x}, b|y) - p(a|\overline{x}, b|\overline{y})\,, \tag{21}$$

for example. It should be clear that Eq. (21) is equivalent to Eq. (20).

Conditional-structure gedankenprobabilities are somewhat different from the inflation DAG kind, in that they reference multiple counterfactual events, such as “What is the probability that Alice would choose to visit the museum *IF* (given that) it’s a rainy day in Maryland *AND* that Alice would choose to go the beach *IF* (given that) it’s sunny in Maryland?”. By contrast, unconditional gedankenprobabilities which live on an inflation DAG reference multiple heterofactual (for lack of a better word) events, such as “What is the probability that Alice-copy-#1 chooses to visit the museum *AND* that it’s raining in Maryland-copy-#1 *AND* that Alice-copy-#2 chooses to go the beach *AND* that it’s sunny in Maryland-copy-#2?”.

Joint counterfactual probabilities are experimentally inaccessible, just the same as joint heterofactual probabilities are. Suppose one could establish both Alice’s propensity for going to the museum when it rains and her propensity for going to the beach when it’s sunny. Even so, neither the joint counterfactual probability nor the joint heterofactual probability can be established from that limited data. For example, the value of the hidden variable $\Lambda{=}\lambda$ may influence Alice’s willingness to get out of bed at all, or determine if she is on-call as an volunteer EMT on a particular day, or λ might encode if Alice is travelling out-of-state. If we could measure $p(a|x, \overline{a}|\overline{x}), ...\}$ we might learn that that Alice’s likelihood of visiting the museum if it rains in Maryland is highly correlated with her likelihood of visiting the beach when it’s sunny in Maryland. Or we might learn that those two counterfactual probabilities are relatively statistically independent. The “hidden-ness” of the latent variable shields the gedankenprobabilities from being determined.

VII. GEDANKENPROBABILITIES AND THE QUANTUM NO-BROADCASTING THEOREM

It is worth emphasizing that certain gedankenprobabilities are strictly classical constructs. If the latent variable in the Bell scenario in Fig. 6 is allowed to be a quantum resource $\mathcal{H}^{d_A \otimes d_B}$, for example, then gedankendistributions such as $P(A|x, A|\overline{x}, ...)$ or $P(A_1, A_2, ...)$ are physically prohibited.

Quantum states are governed by a no-broadcasting theorem [20, 21]: If half the state is sent to Alice and she performs some measurement on it, she fundamentally perturbs the state by measuring it. Post-measurement, that half of the state cannot be “re-sent” to Alice, that she might re-measure it using a different setting. In the inflation DAG picture, a quantum state which was initially available to a single person cannot be distributed both to Alice-copy-#1 and Alice-copy-#2 in the way a classical hidden variable could be, as a consequence of the no-broadcasting theorem.

The meaningless of gedankenprobabilities in the regime of epistemically-restricted general probabilistic theories (GPTs) [21, 41–43], such as quantum theory, means that considerations on inflation DAGs cannot be used to derive quantum causal infeasibility criteria whenever a gendankenprobability presupposes the ability to broadcast a latent variable. When analyzing GPT causal structures one may not assume that a joint probability distribution over observable variables is universally nonnegative if the set of observable variables has simultaneous meaning only under broadcasting of latent variables. This is in direct contrast to Step 1 in Sec. III.

Not every inflation requires broadcasting, however, and not every gedankenprobability is physically prohibited by quantum theory. An inflation DAG is a **non-broadcasting inflation** whenever the children of every individual node in the inflation DAG G' map injectively to the corresponding children of the mapped node in the original DAG G, i.e. $dmap(\mathsf{Ch}_{G'}(X)) \subseteq \mathsf{Ch}_G(dmap(X))$ for all $X \in \mathsf{Nodes}[G']$. Fig. 4 is an example of a non-broadcasting inflation.

Constraints derived from non-broadcasting inflations are valid even in the GPT paradigm. Consequently the inequality in Eq. (7), which was derived from Fig. 4, is therefore a causal infeasibility criterion which holds for the Triangle scenario even what the latent variables are allowed to be quantum resources. As such the GHZ-type distribution per Eq. (11) is found to be infeasible even from the GPT Triangle scenario, recover a result of Henson *et al.* [7]. Indeed, Fig. 5 in Ref. [7] is essentially equivalent to Fig. 4 here.

It might also be possible to derive quantum causal infeasibility criteria if one appropriately modifies Step 1 to generate a different initial set of nonnegativity inequalities. This new set should capture the nonnegativity of only quantum-physically-meaningful marginal probability distributions. From this perspective, a broadcasting inflation DAG is an abstract logical concept, as opposed to a hypothetical physical construct. Indeed, the distributions in a quantum inflation DAG can readily be characterized in terms of the logical broadcasting maps of Coecke and Spekkens [44]. Note that $p(a_1, a_2)$ and other broadcasting-implicit gedankenprobabilities can be *negative* pursuant to a logical broadcasting map.

An analysis along these lines has already been carried out successfully by Chaves *et al.* [17] in the derivation of entropic inequalities for non-classical causal structures. Although Ref. Chaves *et al.* [17] do not invoke inflation DAGs, they do employ conditional-structure, which potentially gives rise to broadcasting-implicit gedankendistributions. Chaves *et al.* [17] take pains to avoid including broadcasting-implicit gedankenentropies in any of their initial inequalities. Nevertheless, deriving *probability* inequalities (as opposed to entropies) has not yet been achieved for non-classical causal structures beyond Bell scenarios.

Our current inflation-DAG method can be employed to derive necessary causal infeasibility criteria for general causal structures, thus generalizing Bell inequalities somewhat. From a quantum foundations perspective, however, generalizing Tsirelson inequalities [15, 45] - the ultimate constraints on what quantum theory makes feasible - is even more desirable. Deriving quantum and/or GPT causal infeasibility for general causal structure is therefore a future research priority.

VIII. TAUTOLOGIES OF THE SUPPLEMENTED EXCLUDED MIDDLE

The process of linear quantifier elimination, while orders of magnitude more computationally amenable than its nonlinear variant, is nevertheless nontrivial. When the number of observable random variables in the inflation DAG is too large it may happen than even advanced linear quantifier elimination algorithms may be too slow for practical application. To this end we have developed a strategy that identifies only one particular class of causal infeasibility criteria, but does so nearly instantly.

In this approach we construct polynomial inequalities corresponding to some tautology of the excluded middle (TEM). In classical logic, the TEM refers to that self-evident truism that every proposition is either true or false, and hence

$$\mathsf{True} \iff \mathsf{Or}\big[A{=}a, A{\neq}a\big] \ . \tag{22}$$

Note that we treat statistical events such as random variables yielding particular outcomes as identically logical propositions. The TEM can also be supplemented with certain "givens", which we take to be known-true propositions on the left-hand-side. These "givens" can then be interspersed throughout the right-rand-side while still yielding a valid tautology. Thus

$$\mathsf{And}\left[\mathbf{a},\mathbf{b}\right] \implies \mathsf{Or}\begin{bmatrix}\mathsf{And}\left[\mathbf{a},c\right],\\ \mathsf{And}\left[\mathbf{b},\overline{c}\right]\end{bmatrix} \tag{23}$$

is an example of a tautology of the *supplemented* excluded middle (T*S*EM). For pedagogical clarity we color and bold the "given" outcomes on both sides of the tautology.

Every TSEM can be converted into a linear inequality by virtue of two connections between classical logical and probability:

1. As the antecedent always implies the consequent, the probability of the antecedent is necessarily less-than-or-equal-to the probability of the consequent. If $j \implies k$ then $p(j) \leq p(k)$.
2. The probability of a disjunction of events is less-than-or-equal-to the sum of the probabilities of the individual events, i.e. $p(j \vee k) = p(j) + p(k) - p(j,k) \leq p(j) + p(k)$.

The inequality which would correspond to the TSEM in Eq. (23) is

$$p(ab) \leq p(ac) + p(b\overline{c})\,. \tag{24}$$

TSEM inequalities can be used as precursors for polynomial inequalities by applying them to inflation DAGs. For example the TSEM inequality

$$p(a_1 b_2) \leq p(a_1 c_1) + p(b_2 \overline{c_1}) \tag{25}$$

applied to our inflation of the Triangle scenario would imply the causal infeasibility criterion

$$p(a)p(b) \leq p(ac) + p(b\overline{c}) \tag{26}$$

pursuant to Fig. 4.

In fact almost every causal infeasibility criterion in this paper is a TSEM inequality! The Triangle scenario inequality Eq. (12) corresponds to the logical TSEM

$$\mathsf{And}\left[\mathbf{a_2},\mathbf{b_2},\mathbf{c_2}\right] \implies \mathsf{Or}\begin{bmatrix}\mathsf{And}\left[\overline{a_1},\overline{b_1},\overline{c_1}\right],\\ \mathsf{And}\left[a_1,\mathbf{b_2},\mathbf{c_2}\right],\\ \mathsf{And}\left[\mathbf{a_2},b_1,\mathbf{c_2}\right],\\ \mathsf{And}\left[\mathbf{a_2},\mathbf{b_2},c_1\right]\end{bmatrix} \tag{27}$$

and then Bell scenario inequality in Eq. (18) follows from the logical TSEM

$$\mathsf{And}\left[\mathbf{a_1},\mathbf{b_1},\mathbf{x_1},\overline{\mathbf{x_2}},\mathbf{y_1},\overline{\mathbf{y_2}}\right] \implies \mathsf{Or}\begin{bmatrix}\mathsf{And}\left[\overline{a_2},\mathbf{b_1},\mathbf{x_1},\overline{\mathbf{x_2}},\mathbf{y_1},\overline{\mathbf{y_2}}\right]\\ \mathsf{And}\left[\mathbf{a_1},\overline{b_2},\mathbf{x_1},\overline{\mathbf{x_2}},\mathbf{y_1},\overline{\mathbf{y_2}}\right]\\ \mathsf{And}\left[a_2,b_2,\mathbf{x_1},\overline{\mathbf{x_2}},\mathbf{y_1},\overline{\mathbf{y_2}}\right]\end{bmatrix} \tag{28}$$

etc.

Generating TSEM inequalities is computationally much easier than performing quantifier elimination. A practical alternative to Sec. III, then, is the four-step process

1. Generate TSEM inequalities on the inflation DAG.
2. Factor the probabilities (same as Step 2 in Sec. III).
3. Map as many probabilities to the original DAG as possible (same as Step 3 in Sec. III).
4. Discard any remaining inequalities which involve unmappable gedankenprobabilities.

Restricting one's search to TSEM inequalities makes deriving causal infeasibility criteria extremely tractable even for large DAGs, but a consequence, however, is that only one class of possible inequalities is being considered.

In order to make a TSEM inequality as compelling as possible one should ensure that every event on the left hand side is also referenced at least once on the right hand side. If one find a TSEM inequality lacking this property then one should tighten the inequalities by deleting any events which occur only on the left hand side. This preserves the tautology while increasing the probability of the left hand side.

We note that the connection between classical propositional logic and linear inequalities has been used previously in the task of causal inference. We reiterate that inequalities resulting from propositional logic, however, are a subset of the inequalities that result from linear quantifier elimination. Consequently, linear quantifier elimination is the preferable tool for deriving inequalities whenever the elimination is computationally tractable. Noteworthy examples of works deriving causal infeasibility criteria via classical logic are Pitowsky [46] and Ghirardi and Marinatto [47]: Eq. (8) here corresponds to Eq. (2-4) in Ref. [46], and Eq. (21) here corresponds to Eq. (30) in Ref. [47].

IX. CONCLUSIONS

Our main contribution is a new way of thinking about classical causal structures, namely the inflation DAG approach. An inflation DAG naturally implies nonlinear hybrid inequalities, i.e. containing gedankenprobabilities, which implicitly constrain the set of feasible distributions per the original causal structure. If desirable, one can further eliminate the gedankenprobabilities via quantifier elimination. Polynomial inequalities can be obtained even through exclusively linear elimination techniques.

These inequalities are necessary but not sufficient causal infeasibility criteria. We've seen that they can be stronger than entropic inequalities sometimes, see Sec. V, yet just how strong they are is still unclear. A distribution might satisfy all our polynomial inequalities and yet not be feasible from the causal structure. Our methods yields tight causal infeasibility criteria for Bell scenarios, but those scenarios are exceptional in that the sets of feasible distributions pursuant to them are convex polytopes.

The most elementary of all causal infeasibility criteria are the conditional independence (CI) relations. Our method explicitly incorporates all marginal independence relations implied by a causal structure. We have found that generic CI relations also appear to be implied by our polynomial inequalities. In future research we hope to clarify the process through which CI relations are manifested as properties of the inflation DAG.

A single causal structure has unlimited potential inflations. Selecting the right inflation from which to derive polynomial inequalities is an interesting challenge. To this end, it would be desirable to understand how particular features of the original causal structure are exposed when different nodes in the DAG are duplicated. By isolating which features are exposed in each inflation we could conceivably quantify the causal inference strength of each inflation. In so doing, we might find that inflated DAGs beyond a certain level of variable duplication need not be considered. The multiplicity beyond which further inflation is irrelevant is presumably related to the maximum degree of those polynomials which tightly characterize a causal scenario. Presently, however, it is not clear how to upper bound either number.

Our method turns the quantum no-cloning theorem [20, 21] on its head by emphasizing that classical latent variables *can*, in-principle, be cloned. This classical cloning possibility motivates the inflation DAG method, and underpins the implied causal infeasibility criteria. We have speculated about generalizing our method to obtain causal infeasibility criteria that constitute necessary constraints even for *quantum* causal scenarios, a common desideratum in recent works [6–8, 17, 28]. It would be enlightening to understand the extent to which our (classical) polynomial inequalities are violated in quantum theory. A variety of techniques exist for estimating the quantum violation of a classical linear inequality [48, 49], but finding the quantum violation of a *polynomial* inequality is a more challenging task [50].

The difference between classical cloning and quantum no-cloning makes the inequalities that result from our consideration to be especially suited for distinguishing the set of quantum-feasible distributions from its subset of classically-feasible distributions. Causal infeasibility criteria that are sensitive to the classical-quantum distinction are precisely the sort of generalizations of the Bell inequalities which are sought after, in order to study the quantum features of generalized causal scenarios. Entropic inequalities have been lacking in this regard [6, 7, 17], and the inflation DAG considerations proposed here constitute a significant alternative strategy.

ACKNOWLEDGMENTS

Research at Perimeter Institute is supported by the Government of Canada through Industry Canada and by the Province of Ontario through the Ministry of Economic Development and Innovation.

[1] J. Pearl, *Causality: Models, Reasoning, and Inference* (Cambridge University Press, 2009).

[2] P. Spirtes, C. Glymour, and R. Scheines, *Causation, Prediction, and Search*, Lecture Notes in Statistics (Springer New York, 2011).

[3] M. Studený, *Probabilistic Conditional Independence Structures*, Information Science and Statistics (Springer London, 2005).

[4] D. Koller, *Probabilistic Graphical Models : Principles and Techniques* (MIT Press, Cambridge, MA, 2009).

[5] 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).

[6] T. Fritz, "Beyond Bell's theorem: correlation scenarios," New J. Phys. **14**, 103001 (2012).

[7] J. Henson, R. Lal, and M. F. Pusey, "Theory-independent limits on correlations from generalized bayesian networks," New J. Phys. **16**, 113043 (2014).

[8] T. Fritz, "Beyond Bell's Theorem II: Scenarios with Arbitrary Causal Structure," Comm. Math. Phys. **341**, 391 (2015).

[9] T. Fritz and R. Chaves, "Entropic inequalities and marginal problems," IEEE Trans. Info. Theo. **59**, 803 (2013).

[10] R. Chaves, L. Luft, and D. Gross, "Causal structures from entropic information: geometry and novel scenarios," New J. Phys. **16**, 043001 (2014).

[11] R. Chaves, L. Luft, T. O. Maciel, D. Gross, D. Janzing, and B. Schölkopf, "Inferring latent structures via information inequalities," in *Proc. of the 30th Conference on Uncertainty in Artificial Intelligence* (AUAI, 2014) pp. 112–121.

[12] J. S. Bell, "On the Problem of Hidden Variables in Quantum Mechanics," Rev. Mod. Phys. **38**, 447 (1966).

[13] J.-D. Bancal, C. Branciard, N. Brunner, N. Gisin, and Y.-C. Liang, "A framework for the study of symmetric full-correlation Bell-like inequalities," J. Phys. A **45**, 125301 (2012).

[14] V. Scarani, "The device-independent outlook on quantum physics," Acta Physica Slovaca **62**, 347 (2012).

[15] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, "Bell nonlocality," Rev. Mod. Phys. **86**, 419 (2014).

[16] T. Fritz, "Polyhedral duality in Bell scenarios with two binary observables," J. Math. Phys. **53**, 072202 (2012), 10.1063/1.4734586.

[17] R. Chaves, C. Majenz, and D. Gross, "Information–theoretic implications of quantum causal structures," Nat. Comm. **6**, 5766 (2015).

[18] C. Branciard, D. Rosset, N. Gisin, and S. Pironio, "Bilocal versus nonbilocal correlations in entanglement-swapping experiments," Phys. Rev. A **85**, 032119 (2012).

[19] B. Steudel and N. Ay, "Information-theoretic inference of common ancestors," arXiv:1010.5720 (2010).

[20] H. Barnum, C. M. Caves, C. A. Fuchs, R. Jozsa, and B. Schumacher, "Noncommuting mixed states cannot be broadcast," Phys. Rev. Lett. **76**, 2818 (1996).

[21] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, "Cloning and Broadcasting in Generic Probabilistic Theories," quant-ph/0611295 (2006).

[22] C. Barrett, P. Fontaine, and C. Tinelli, "The Satisfiability Modulo Theories Library (SMT-LIB)," `www.SMT-LIB.org` (2016).

[23] R. Chaves, "Polynomial bell inequalities," Phys. Rev. Lett. **116**, 010402 (2016).

[24] C. Jones, E. C. Kerrigan, and J. Maciejowski, *Equality set projection: A new algorithm for the projection of polytopes in halfspace representation*, Tech. Rep. (Cambridge University Engineering Dept, 2004).

[25] C. Jones, *Polyhedral Tools for Control*, Ph.D. thesis, University of Cambridge (2005).

[26] C. N. Jones, E. C. Kerrigan, and J. M. Maciejowski, "On polyhedral projection and parametric programming," J. Optimiz. Theo. Applic. **138**, 207 (2008).

[27] D. Avis, "Polytopes — combinatorics and computation," (Birkhäuser Basel, Basel, 2000) Chap. A Revised Implementation of the Reverse Search Vertex Enumeration Algorithm, pp. 177–198.

[28] R. Chaves and C. Budroni, "Entropic nonsignalling correlations," arXiv:1601.07555 (2015).

[29] A. Tavakoli, P. Skrzypczyk, D. Cavalcanti, and A. Acín, "Nonlocal correlations in the star-network configuration," Phys. Rev. A **90**, 062109 (2014).

[30] D. Rosset, C. Branciard, T. J. Barnea, G. Pütz, N. Brunner, and N. Gisin, "Nonlinear bell inequalities tailored for quantum networks," Phys. Rev. Lett. **116**, 010403 (2016).

[31] A. Tavakoli, "Bell-type inequalities for arbitrary noncyclic networks," Phys. Rev. A **93**, 030101 (2016).

[32] C. Kang and J. Tian, "Polynomial constraints in causal bayesian networks," in *Proc. of the 23rd Conference on Uncertainty in Artificial Intelligence* (AUAI, 2007) pp. 200–208.

[33] G. V. Steeg and A. Galstyan, "A sequence of relaxations constraining hidden variable models," in *Proc. 27th Conference on Uncertainty in Artificial Intelligence* (AUAI, 2011) pp. 717–726.

[34] B. W. Schumacher, "Information and quantum nonseparability," Phys. Rev. A **44**, 7047 (1991).

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

[36] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, "Proposed experiment to test local hidden-variable theories," Phys. Rev. Lett. **23**, 880 (1969).

[37] J.-D. Bancal, *On the Device-Independent Approach to Quantum Physics* (Springer International Publishing, 2014).

[38] J. S. Bell, "On the Einstein-Podolsky-Rosen paradox," Physics **1**, 195 (1964).

[39] J. M. Donohue and E. Wolfe, "Identifying nonconvexity in the sets of limited-dimension quantum correlations," Phys. Rev. A **92**, 062120 (2015).

[40] J. F. Clauser and M. A. Horne, "Experimental consequences of objective local theories," Phys. Rev. D **10**, 526 (1974).

[41] R. W. Spekkens, "Evidence for the epistemic view of quantum states: A toy theory," Phys. Rev. A **75**, 032110 (2007).

[42] H. Barnum and A. Wilce, "Post-Classical Probability Theory," arXiv:1205.3833 (2012).

[43] P. Janotta and H. Hinrichsen, "Generalized probability theories: what determines the structure of quantum theory?" J. Phys. A **47**, 323001 (2014).

[44] B. Coecke and R. W. Spekkens, "Picturing classical and quantum bayesian inference," Synthese **186**, 651 (2011).

[45] B. S. Cirel'son, "Quantum generalizations of Bell's inequality," Lett. Math. Phys. **4**, 93 (1980).

[46] I. Pitowsky, *Quantum Probability - Quantum Logic*, Lecture Notes in Physics, Berlin Springer Verlag, Vol. 321 (Springer Berlin Heidelberg, 1989).

[47] G. Ghirardi and L. Marinatto, "Proofs of nonlocality without inequalities revisited," Phys. Lett. A **372**, 1982 (2008).

[48] M. Navascués, S. Pironio, and A. Acín, "A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations," New J. Phys. **10**, 073013 (2008).

[49] K. F. Pál and T. Vértesi, "Quantum Bounds on Bell Inequalities," Phys. Rev. A **79**, 022120 (2009).

[50] S. Pironio, M. Navascués, and A. Acín, "Convergent relaxations of polynomial optimization problems with noncommuting variables," SIAM J. Optim. **20**, 2157 (2010).

[51] E. Balas, "Projection with a minimal system of inequalities," Comp. Optimiz. Applic. **10**, 189 (1998).

[52] A. Fordan, *Projection in Constraint Logic Programming* (Ios Press, 1999).

Appendix A: Further Polytope Projection Algorithms

The Equality Set Projection (ESP) algorithm [24, 25] is ideal for handling inflation DAGs, because its computational complexity scales only according to the facet count of the final projection. Our use of larger-and-larger inflation DAGs to obtain causal infeasibility criteria on the same underlying original DAG means that while the complexity of the starting polytope is unbounded, the complexity of the projection is finite. Practically, this suggests that the ESP algorithm could parse the implications due to a very large inflation DAGs efficiently. Formally, ESP should require minimal computational overhead to consider a larger inflation DAG relative to considering a much smaller inflation DAG, when the *implications* of the small and large inflations are similar. By contrast, the computation complexity of Fourier-Motzkin (FM) elimination algorithm scales with the number of quantifiers being eliminated. The number of gedankenprobabilities requiring elimination is exponentially related to the number of variables in the inflation DAG. The FM algorithm, therefore, is utterly impractical very for large inflation DAGs.

Another positive feature of the ESP algorithm is that it commences outputting quantifier-free inequalities immediately, and terminates upon deriving the complete set of inequalities. By contrast, FM works by eliminating one quantifier at a time. Terminating the ESP algorithm before it reaches completion would result in an incomplete list of inequalities. Even an incomplete list is valuable, though, since the causal infeasibility criteria we are deriving are anyways necessary but not sufficient.

Vertex projection (VP) algorithms are another computational tool which may be used to assist in linear quantifier elimination [27]. VP works by first enumerating the vertices of the initial polytope (H-rep to V-rep), projecting the vertices, and then converting back to inequalities (V-rep to H-rep). For generic high-dimensional polytopes, the operation of converting from a representation in terms of halfspaces to one in terms of extremal-vertices representations can be computationally costly (high-d H-rep to V-rep). Starting from a vertex representation in a high dimensional space, however, one can immediately determine the vertex representation of the polytope's projection in a lower dimensional space. The projection is along the coordinate axes, so one just "discards" the coordinate of the eliminated quantifier. To obtain the inequalities which characterize the projected polytope one then applies a convex hull algorithm to the projected vertices (low-d V-rep to H-rep).

For probability distributions, however, the extremal vertices are precisely the deterministic possibilities. Since the extremal vertices of the initial polytope are easily enumerated, it is possible to avoid the high-d V-rep to H-rep step entirely. There is a one-to-one correspondence between the inflation-DAG's initial generating inequalities and its initial extreme observable probability distributions. We used this V-rep to H-rep technique to project the initial polytope implied by Fig. 3 (Step 1) to an intermediate 23-dimensional polytope, where each of the 23 remaining can be mapped the original DAG. Only then did we apply the transformations of factorization and mappings (Steps. 2,3) to convert those linear inequalities to polynomial inequalities pertaining to the original DAG. We found that the V-rep to H-rep technique, using *lrs* [lrs], was orders-of-magnitude faster than FM elimination at obtaining the same result.

Yet another technique is also possible. Suppose the initial polytope is given by $\left\{\vec{x},\vec{y}\,|\hat{A}.\vec{x}+\hat{B}.\vec{y}\geq \boldsymbol{c}\right\}$, where y are the quantifiers. If we can find any completely nonnegative vector $\boldsymbol{w}$ such that $\boldsymbol{w}.\hat{B}=\vec{0}$ then we automatically establish the quantifier-free inequality $\boldsymbol{w}.\hat{A}.\vec{x}\geq \boldsymbol{w}.\boldsymbol{c}$. Solving for "random" nonnegative vectors $\boldsymbol{w}$ is easy; solving for all possible solutions is rather more difficult. Balas [51] refined this method so that each extremal construction of $\boldsymbol{w}$ corresponds to an irredundant inequality in the H-rep description of the projected polytope. Nevertheless, even without utilizing the full projection cone, this technique can be used to rapidly obtain a few quantifier-free inequalities.

Appendix B: Optimized Algorithm for Recognizing Redundant Inequalities

When performing Fourier-Motzkin linear quantifier elimination one must periodically filter out redundant inequalities from the set of linear inequalities. Equivalently, the means identifying redundant halfspace constraints in the description of the polytope. An individual constraint in a set is redundant if it is implied by the other constraints.

An individual linear inequality is redundant if and only if it is a *positive* linear combination of the others [Thm. 5.8 in 52][7]. This is related to the V-rep characterization of polyhedral cones: If a cone is defined such that $W_{\hat{M}}:=\left\{\vec{x}\,|\exists_{\boldsymbol{v}\geq\boldsymbol{0}}:\,\hat{M}.\boldsymbol{v}=\vec{x}\right\}$ then $\vec{b}\in W_{\hat{M}}$ if and only if the linear system of equations $\hat{M}.\boldsymbol{v}=\vec{b}$ has a solution such that all the elements of $\boldsymbol{v}$ are nonnegative. Thus, the computational tool required is one which accepts as input the matrix $\hat{M}$ and the column vector $\vec{b}$ and returns $\vec{b}\in W_{\hat{M}}$ as True or False.

Below, we present two possible *Mathematica*™ implementations which assess if a given column $\vec{b}$ can be expressed as a positive linear combination of the columns of $\hat{M}$. The former function is easy to understand, but the latter utilizes efficient low-level code and *Mathematica*™'s internal error-handling to rapidly recognize infeasible linear programs.

```
PositiveLinearSolveTest[M_?MatrixQ, b_]:=With[{vars = Thread[Subscript[x, Dimensions[M][[2]]]]},
     Resolve[Exists[Evaluate[vars], AllTrue[vars, NonNegative],
     And@@Thread[M.vars == Flatten[b]]]]];
```
or
```
PositiveLinearSolveTest[M_?MatrixQ, b_]/;Dimensions[b]==={Length[M], 1}:=
Module[{rowcount, columncount, fakeobjective, zeroescolumn},
 {rowcount, columncount} = Dimensions[M];
 fakeobjective = SparseArray[{}, {columncount}, 0.0]; zeroescolumn = SparseArray[{}, {rowcount, 1}];
 Internal`HandlerBlock[{Message, Switch[#1, Hold[Message[LinearProgramming::lpsnf, ___], _], Throw[False]]&},
     Quiet[Catch[
       LinearProgramming[fakeobjective, M, Join[b, zeroescolumn, 2], Method -> Simplex]; True
     ], {LinearProgramming::lpsnf}]]];
```

To illustrate examples of a when a positive solution to the linear system exists and when it does not, consider the following two examples:.

$$\texttt{PositiveLinearSolveTest}[\begin{pmatrix}1&0&1\\0&1&-1\end{pmatrix},\begin{pmatrix}1\\-2\end{pmatrix}]\ \texttt{==}\ \texttt{False}$$

$$\texttt{PositiveLinearSolveTest}[\begin{pmatrix}1&0&1\\0&1&-2\end{pmatrix},\begin{pmatrix}1\\-1\end{pmatrix}]\ \texttt{==}\ \texttt{True}$$

If $\hat{A}$ is the matrix who's rows are nonnegativity inequalities, then the following test determines if row n is redundant.

```
RedundantRowQ[A_?MatrixQ, n_Integer]:=PositiveLinearSolveTest@@Reverse[Transpose/@TakeDrop[A, {n}]].
```

Note that a `True` response from `RedundantRowQ` indicates that the row n is redundant.

[7] The "if" is obvious. The "only if" is a consequence of Farka's lemma [52].

Appendix C: Recognizing observationally equivalent DAGs

One expects that an edge $A \to B$ can be added to DAG G while leaving G observationally invariant if the new connection does not introduce any new information about observable variables to B. We can formalize this notion in the language of sufficient statistics. To do so, however, a few background definitions are in order.

Perfectly Predictable: The random variable X is perfectly predictable from a set of variables $\boldsymbol{Z}$, hereafter $\boldsymbol{Z} \vDash X$, if X can be completely inferred from knowledge of $\boldsymbol{Z}$ alone. In a deterministic DAG, for example, every non-root node is perfectly predictable given its parents, $\mathsf{pa}\left[X\right] \vDash X$. Indeed, in a deterministic DAG the node X is perfectly predictable from $\boldsymbol{Z}$ if X is a deterministic descendant of $\boldsymbol{Z}$. Operationally, X is a deterministic descendant of Z if the intersection of [the ancestors of X] with [the non-ancestors of Z] is a subset of [the descendants of Z]. Happily though, perfectly predictability can be extrapolated from a causal structure with minimal effort: $\boldsymbol{Z} \vDash X$ if every directed path to X from any root node is blocked by $\boldsymbol{Z}$.

Markov Blanket: The Markov Blanket for a set of nodes $\boldsymbol{V}$, hereafter $\mathsf{MB}\left[\boldsymbol{V}\right]$, is the set of all of $\boldsymbol{V}$'s children, parents, and co-parents. The Markov Blanket is so defined because the nodes in $\boldsymbol{V}$ are conditionally independent of *everything* given $\mathsf{MB}\left[\boldsymbol{V}\right]$. If the random variables in the Markov Blanket $\mathsf{MB}\left[\boldsymbol{V}\right]$ are known, then information about nodes inside $\boldsymbol{V}$ has no bearing on nodes outside the Markov Blanket and vice versa.

Markov Partition: New! I made this up Nov 24. Useful do you think? A set of variables $\boldsymbol{Z}$ is a Markov Partition for a pair of random variables X and Y, hereafter $X{\dashv}\boldsymbol{Z}{\vdash}Y$, if the pair are conditionally independent of eachother given *any superset* of $\boldsymbol{Z}$. Operationally, this means that X and Y are d-separated by every superset of $\boldsymbol{Z}$. Equivalently, $X{\dashv}\boldsymbol{Z}{\vdash}Y$ if $\mathsf{MB}\left[\boldsymbol{V}\right] \subseteq \boldsymbol{Z}$ and $X \in \boldsymbol{V}$ while $Y \notin \boldsymbol{V}$, or if $\mathsf{MB}\left[\boldsymbol{V}\right] \subseteq \boldsymbol{Z}$ and $Y \in \boldsymbol{V}$ while $X \notin \boldsymbol{V}$. Happily though, Markov Partitions can be extrapolated from a causal structure with minimal effort: $X{\dashv}\boldsymbol{Z}{\vdash}Y$ if and only if X and Y would be in *disconnected components* under the deletion of all edges initiation from $\boldsymbol{Z}$.

Sufficient Statistic: A set of nodes $\boldsymbol{Z}$ is a sufficient statistic for A relative to X, hereafter $\boldsymbol{Z} \vdash A|X$, if and only if all inferences about X which can be made given knowledge of A are also inferable *without* knowing A but with knowing $\boldsymbol{Z}$ instead. In other words, learning A can never teach anything new about X if $\boldsymbol{Z}$ is already known. If $X = A$, then the *only way* $\boldsymbol{Z}$ can stand in for A when making inferences about A is if A is perfectly predicable given $\boldsymbol{Z}$, i.e. $\boldsymbol{Z} \vdash A|A \iff \boldsymbol{Z} \vDash A$. If $A \neq B$ then there are four and only four?) ways that $\boldsymbol{Z} \vdash A|X$ can be implied by a DAG: If $\boldsymbol{Z} \vDash A$, if $\boldsymbol{Z} \vDash X$, if $\mathsf{MB}\left[\boldsymbol{V}\right] \subseteq \boldsymbol{Z}$ and $A \in \boldsymbol{V}$ while $X \notin \boldsymbol{V}$, or if $\mathsf{MB}\left[\boldsymbol{V}\right] \subseteq \boldsymbol{Z}$ and $X \in \boldsymbol{V}$ while $A \notin \boldsymbol{V}$. Alternatively: If $A \neq B$ then there are THREE ways that $\boldsymbol{Z} \vdash A|X$ can be implied by a DAG: If $\boldsymbol{Z} \vDash A$, if $\boldsymbol{Z} \vDash X$, and if $A{\dashv}\boldsymbol{Z}{\vdash}X$.

Theorem 1. *An edge $A \to B$ can be added to G without observational impact if $\mathsf{pa}\left[B\right]$ are a sufficient statistic for A relative to all observable nodes, i.e. $\forall_{observable\ X} : \mathsf{pa}\left[B\right] \vdash A|X$.*
In particular, the edge $A \to B$ can always be added whenever $\mathsf{pa}\left[B\right] \vDash A$, including, but not limited to, the instance $\mathsf{pa}\left[A\right] \subseteq \mathsf{pa}\left[B\right]$.
Furthermore, the edge $\Lambda \to B$ can be also always be added whenever Λ is latent and $\mathsf{MB}\left[\Lambda\right] \subseteq \mathsf{pa}\left[B\right]$.

We can also define an analogous condition for when an edge can be removed from a DAG without impacting it observationally.

Corollary 1.1. *An edge $A \to B$ can be dropped from G to form G' such that G and G' are observationally equivalent if ~~and only if~~ the edge $A \to B$ can be added (back) to G' while leaving G' observationally invariant per Theorem 1.*

On the subject of adding observationally-invariant edges, it is important to recognize when latent variables can be introduced (or dropped) without observational impact.

Theorem 2. *A (root) latent variable Λ can be removed from G without observational impact if Λ has only one child node and no co-parents (Λ is "useless"), or if Λ's children are also all children of another single latent variable (Λ is "redundant"). Conversely, a new root latent variable Λ can be introduced along with various outgoing edges, without observational impact, if Λ would be useless or redundant*

Naturally, two causal structures are observationally equivalent if one can be transformed into the other without observational impact, via Theorems 1 and 2. Some examples of observationally equivalent scenarios, and the steps which interconvert them, are given in Fig. 9.

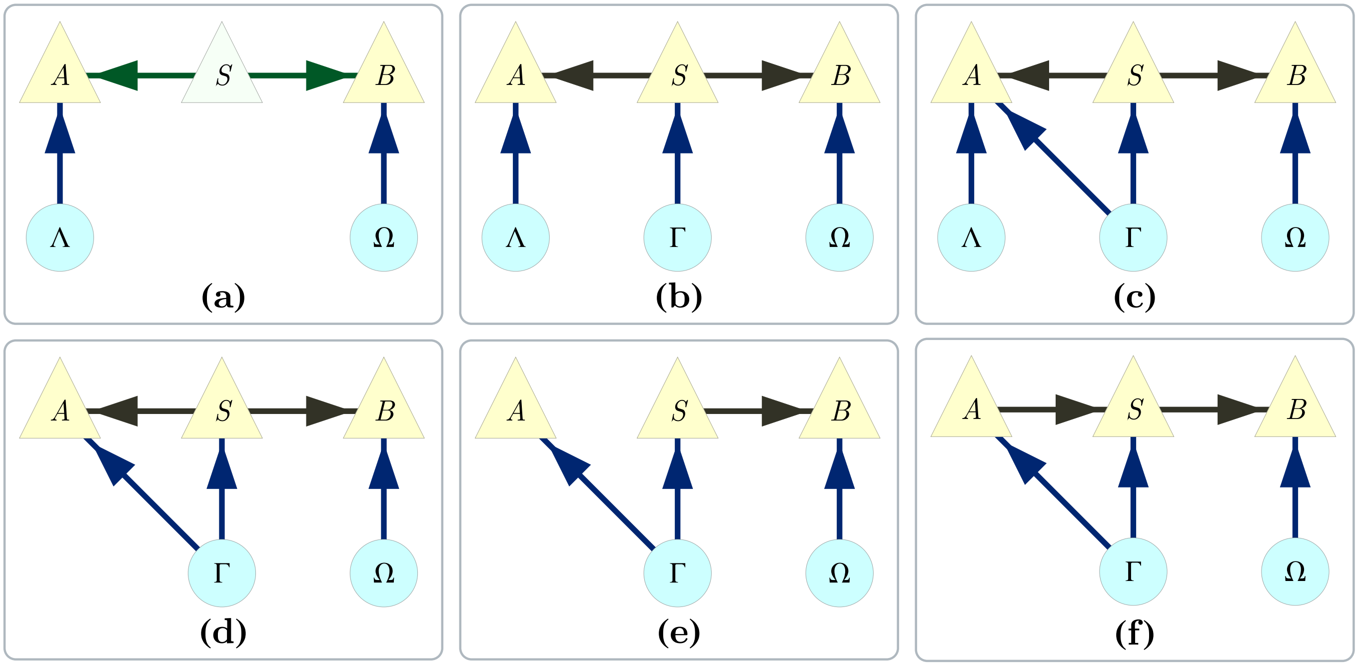

FIG. 9. A set of observational equivalent causal structures. The reasons the changes are observational invariant are as follows:
(a)∼(b) because Γ is useless in (b), and as such Γ can be dropped from (b) per Theorem 2.
(b)∼(c) because $\mathsf{MB}[\Gamma] \subseteq \mathsf{pa}[A]$ in (b), and as such $\Gamma \to A$ can be added to (b) per Theorem 1.
(c)∼(d) because Λ is redundant to Γ in (c), and as such Λ can be dropped from (c) per Theorem 2.
(d)∼(e) because $\mathsf{pa}[S] \subseteq \mathsf{pa}[A]$ in (e), and as such $S \to A$ can be added to (e) per Theorem 1.
(e)∼(f) because $\mathsf{pa}[A] \subseteq \mathsf{pa}[S]$ in (e), and as such $A \to S$ can be added to (e) per Theorem 1.

Appendix D: Tobias's Original 7 Inequalities

"I present several inequalities... together with a method of proof which has a combinatorial flavour. No quantum violations of any of these inequalities has been found to date."

Theorem 3. *The following inequalities hold for all classical correlations in the triangle scenario:*

(a) $p(a)p(c) \le p(ab) + p(\bar{b}c)$

(b) $p(ab\bar{c})p(a\bar{b}c)p(\bar{a}bc) \le p(abc) + p(\bar{a}\bar{b})p(ab\bar{c})p(a) + p(\bar{a}\bar{c})p(a\bar{b}c)p(c) + p(\bar{b}\bar{c})p(\bar{a}bc)p(b)$

(c) $p(ab\bar{c})p(a\bar{b}c)p(\bar{a}bc) \le p(abc)^2 + 2\big(p(\bar{a}\bar{b})p(ab\bar{c}) + p(\bar{a}\bar{c})p(a\bar{b}c) + p(\bar{b}\bar{c})p(\bar{a}bc)\big)$

(d) $p(abc)^2 p(\bar{a}\bar{b}\bar{c}) \le p(ab\bar{c})p(a\bar{b}c)p(\bar{a}bc) + \big(2p(abc) + p(\bar{a}\bar{b}\bar{c})\big)\big(1 - p(abc) - p(\bar{a}\bar{b}\bar{c})\big)$

(e) $p(abc)^2 p(\bar{a}\bar{b}\bar{c}) \le p(abc)^3 + \big(2p(abc) + p(\bar{a}\bar{b}\bar{c})\big)\big(1 - p(abc) - p(\bar{a}\bar{b}\bar{c})\big)$

(f) $p(a)p(b)p(c) \le p(\bar{a}\bar{b}\bar{c}) + p(ab)p(c) + p(ac)p(b) + p(bc)p(a)$

(g) $p(a)p(b)p(c) \le p(\bar{a}\bar{b}\bar{c})^2 + 2\Big(p(ab)p(c) + p(ac)p(b) + p(bc)p(a)\Big)$

It is quite likely that some of these inequalities are dominated by the others, but I do not know for sure whether any of them are actually redundant."

Note that Eq. (13) implies inequalities (a), (b), and (f). I haven't checked the others yet. ~EW