

The Inflation DAG Technique for Causal Inference with Hidden Variables

Elie Wolfe,^{1,*} Robert W. Spekkens,^{1,†} and Tobias Fritz^{1,2,‡}

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: April 26, 2016)

The fundamental problem of causal inference is to infer from a given probability distribution over observed variables, what causal structures, possibly incorporating hidden variables, could have given rise to that distribution. Given some candidate causal structure, it is therefore valuable to derive infeasibility criteria, such that the hypothesis is not a feasible causal explanation whenever the observed distribution violates an infeasibility criterion. The problem of causal inference via infeasibility criteria comes up in many fields. Special infeasibility criteria are Bell inequalities (which distinguish non-classical from classical distributions) and Tsirelson inequalities (which distinguish quantum from post-quantum distributions), and Pearl’s instrumental inequality. All of these are limited to very specific causal structures. Analogues of such inequalities for more-general causal structures, i.e., necessary criteria for either classical or quantum distributions to be realizable from the structure, are highly sought after.

We here introduce a technique for deriving such infeasibility criteria, applicable to any causal structure. It consists of first *inflating* the causal structure and then translating weak constraints on the inflated structure into stronger constraints on the original structure. Moreover, we show how our technique can be tuned to yield either classical criteria (i.e., that may have quantum violations), or post-classical criteria (i.e., that hold even in the context of general probability theories), depending on whether or not the inflation implicitly broadcasts the value of a hidden variable. Concretely, we derive polynomial inequalities for the so-called Triangle scenario, and we show how all Bell inequalities also follow from our method. Furthermore, given both a causal structure and a specific probability distribution, our technique can be used to efficiently witness their inconsistency, even absent explicit inequalities. The inflation technique is therefore both relevant and practical for general causal inference tasks with hidden variables.

* ewolfe@perimeterinstitute.ca

† rspekkens@perimeterinstitute.ca

‡ tobias.fritz@mis.mpg.de

I. INTRODUCTION

Given some hypothesis of causal structure, it is desirable to determine **infeasibility criteria**, i.e., observable constraints such that their violation implies the invalidity of the hypothesis as an explanation for observational data. Causal infeasibility criteria are used in a wide variety of statistics application, from sussing out biological pathways to enabling machine learning [1–4]. **ADD SENTENCES ABOUT HOW OUR WORK CONTRIBUTES TO GENERAL CAUSAL INFERENCE TASKS.** The foundational role of causal structure in quantum information theory has only recently been appreciated [5–8].

In contexts other than quantum theory, the latent nodes in causal structures are generally taken to represent hidden variables. This is not fully general, however, so we apply the retronym¹ “classical”, as in classical causal structure and classical causal inference. The classical distributions of a given causal structure are defined as those which arise from it while restricting the latent nodes to be arbitrary (classical) random variables. Quantum distributions, by contrast, are those which are realizable if the latent nodes in the causal structure are allowed to be quantum systems. We hereafter take all causal structures and probability distributions to be classical, except where explicitly stated otherwise.

From a physics perspective, therefore, tightly characterizing the set of observable probability distributions realizable from a causal structure is critical, in order to recognize and exploit the existence of distributions that can be realized quantumly but not classically. Few techniques are known for bounding this set of distributions which are simultaneously practical and applicable to general causal structures. Celebrated examples include the use of conditional independence relations (easy) [1–4] and entropic inequalities (more advanced) [10–12]. In the presence of hidden variables, these criteria only rarely provide a tight characterization, and frequently fail to witness the non-classicality of quantum distributions.

Distinguishing quantum from classical correlations has historically been achieved through the use of Bell inequalities [13–17]. Bell inequalities, however, are limited to very special causal scenarios involving *only one* latent common cause variable, i.e. Bell scenarios. A Bell scenario is also very special in that its realizable distributions admit characterization by a finite set of linear inequalities (after conditioning on the setting variables), i.e. its realizable distributions comprise a convex polytope [14, 18]. Entirely new techniques, therefore, are required to derive quantum-sensitive infeasibility criteria for more general causal scenarios [6–8].

To this end, we here introduce a new technique, applicable to any causal structure, for deriving infeasibility criteria. This technique allows for, but is not limited to, the derivation of polynomial inequalities. These criteria are generally based on the *broadcasting* of the values of a hidden variable, i.e. the assumption that its value can be copied and broadcast at will. The no-broadcasting theorem from quantum theory shows that this is not valid in the non-classical case, and from our perspective this is the reason for the existence of quantum violations of Bell inequalities. Moreover, our technique can also be applied in order to derive criteria that must be satisfied for all distributions that can be generated with latent nodes that are states in quantum theory or any other general probabilistic theory, simply by not assuming the possibility of broadcasting.

II. CAUSAL MODELS AND CAUSAL INFERENCE

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 amount of sleep S . The logical proposition, or **event**, $X=x, S=s$ should be understood as “the student scores x on the exam with 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 particular events. Thus $P(X, Y)$ is the multivariate probability distribution over the random variables $\{XY\}$, and $p(x, y)$ denotes the joint probability of the two events $X=x$ and $Y=y$. We often omit the comma and just write $P(XY)$ and $p(xy)$, respectively.

A causal model consists of a pair of objects: a causal structure and a set of parameters. The causal structure is a directed acyclic graph (DAG). Recall that a DAG G consists of a set of nodes and directed edges (i.e., ordered pairs of nodes), which we denote by $\text{Nodes}[G]$ and $\text{Edges}[G]$ respectively. Physically, each node in the DAG corresponds to a localized system and a directed edge between two nodes corresponds to there being a direct causal influence from one system to the other. The parameters specify, for each root node, how the system corresponding to that node is prepared, and for each node that has parents in the DAG, the precise manner in which it causally depends on its

¹ Retronym (noun): a modification of an original term to distinguish it from a later development [9].

parents. If a DAG is denoted G and the parameter values thereon are denoted $\mathcal{P}(G)$, then the corresponding causal model is the pair $C = (G, \mathcal{P}(G))$. We shall also make use of the notation $\text{DAG}[C]$ for G and $\text{ParamVals}[C]$ for $\mathcal{P}(G)$.

Our terminology for the causal relations between the nodes in a DAG is the standard one. The parents of a node X in a given graph G are defined as those nodes which have directed edges originating at them and terminating at X , i.e. $\text{Pa}_G(X) = \{Y \mid Y \rightarrow X\}$. 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. $\text{Ch}_G(X) = \{Y \mid X \rightarrow Y\}$. If U is a set of nodes, then we put $\text{Pa}_G(U) := \bigcup_{X \in U} \text{Pa}_G(X)$ and $\text{Ch}_G(U) := \bigcup_{X \in U} \text{Ch}_G(X)$. The **ancestors** of a set of nodes U , denoted $\text{An}_G(U)$, are defined as those nodes which have a directed *path* to some node in U , including the nodes in U themselves. Equivalently (dropping the G subscript), $\text{An}(U) := \bigcup_{n \in \mathbb{N}} \text{Pa}^n(U)$, where $\text{Pa}^n(U)$ is inductively defined via $\text{Pa}^0(U) := U$ and $\text{Pa}^{n+1}(U) := \text{Pa}(\text{Pa}^n(U))$.

The subgraph of G induced by restricting attention to the set of nodes V will be denoted $\text{SubDAG}_G[V]$. It consists of the nodes V and the edges between pairs of nodes in V per the original graph. Of special importance to us is the **ancestral subgraph** of V , denoted $\text{AnSubDAG}_G[V]$, which is the minimal subgraph containing the full ancestry of V , $\text{AnSubDAG}_G[V] := \text{SubDAG}_G[\text{An}_G(V)]$.

In a causal model, the localized systems are random variables, and the parameters are conditional probabilities². Specifically, for each node, the model specifies a conditional probability distribution over the values of the random variable associated to that node, given the values of the variables associated to its parents in the DAG. (In the case of root nodes, the parents are the null set and the conditional probability distribution is simply a probability distribution.) Denoting the conditional probability distribution associated to a node A in the causal model C by $P_C(A|\text{Pa}_G(A))$, we have

$$\mathcal{P}(G) \equiv \{P_C(A|\text{Pa}_G(A)) : A \in \text{Nodes}[G]\}. \quad (1)$$

As is standard, we assume that the joint distribution defined by the causal model is $P_C(\text{Nodes}[G]) = \prod_{A \in \text{Nodes}[G]} P_C(A|\text{Pa}_G(A))$ (this is called the Markov condition). The subset of nodes of G that are observed is denoted $\text{ObservedNodes}[G]$ and the rest are denoted $\text{LatentNodes}[G]$. The joint distribution over the observed variables is obtained from the joint distribution over all variables by marginalization over the latent variables $P_C(\text{ObservedNodes}[G]) = \sum_{\{X: X \in \text{LatentNodes}[G]\}} P_C(\text{Nodes}[G])$.

In addition to the notion of a causal model, we define the notion of a causal hypothesis, which is a set of causal models consistent with a single DAG. Specifically, if G is a DAG and S is a set of parameter values consistent with G (i.e. a set of possibilities for the conditional probability distributions at each node of G), then G and S define a causal hypothesis, which we denote by $H_{G,S}$, and which we define formally as the following set of causal models:

$$H_{G,S} \equiv \{C \mid \text{DAG}[C] = G, \text{ParamVals}[C] \in S\}. \quad (2)$$

We denote the set of *all* parameter values consistent with G by S_{full} , so that $H_{G,S_{\text{full}}}$ is simply the causal hypothesis that the DAG is G , that is, the hypothesis that there exists *some* parameter values supplementing G such that the observed data can be explained by the resulting causal model.

With these notions in hand, we can give a broad definition of causal inference. An algorithm for causal inference is one that takes two inputs:

- a specification of *observational data*
- a specification of a *causal hypothesis*

and that outputs a verdict about their compatibility.

The most informative sort of observational data is a specification of the joint distributions over all of the observed variables. The observational data in real-world applications necessarily falls short of this ideal because one only ever obtains a *finite sample* of the joint distribution. Moreover, even if one sets aside finite-sample concerns and imagines the data as infinite-run statistics, there are other ways in which the observational data for a causal inference problem might be less informative than a specification of the joint distribution. For instance, one might merely have a description of the conditional independence relations that hold among the observed variables. Or one might merely have a specification of the marginals of the joint distribution for certain subsets of the observed variables, in which

² This is the definitio of a *classical* causal model, which will be the primary focus of this article. Nonetheless, there is a quantum generalization of the notion of a causal model that will feature in our discussions at the end of this article. In a *quantum* causal model [19], the localized systems are types of quantum systems (specified by Hilbert space dimensionality), and the parameters are quantum operations. Specifically, for each root node, the model specifies a unit-trace positive operator on the Hilbert space for that system and for each non-root node, the model specifies a trace-preserving completely-positive linear map from the operators on the Hilbert space of the parents of the system to the operators on the Hilbert space of the system. Certain nodes in a quantum causal model may be considered classical, in which case all states and operations must be diagonal in some fixed basis on that system. One can also generalize the notion of a causal model further still to the case of generalized probabilistic theories, as was done in Ref. [7].

case the causal inference problem corresponds to a version of the marginals problem, but where the joint distribution is constrained by the causal hypothesis. Indeed, this latter sort of causally-constrained marginals problem will be relevant for our inflation DAG technique.

The most fine-grained sort of causal hypothesis is one that specifies not only a DAG but all of the parameters supplementing the DAG. The causal inference problem that is perhaps most widely studied is one wherein the causal hypothesis just specifies a DAG. Between these two cases are causal inference problems that explore more refined causal hypotheses, incorporating constraints on how certain nodes in the DAG functionally depend on their parents. An example is the assumption of an additive noise model for certain causal influences: if an observed variable Y has an observed variable X as a parent and also a latent variable U , then the noise is deemed additive if $Y = \alpha X + \beta U$ for some scalars α and β [provide references]. More general constraints on the causal dependences in the model have also been explored [provide references]. (Note that constraints on the probability distributions over the latent variables can be understood as a kind of constraint on the causal dependences.) Our inflation DAG technique will require consideration of a novel and unusual sort of constraint on the causal dependences of the model.

Finally, the verdict about the compatibility of the observational input and the causal hypothesis could come in different forms. The simplest possibility is a specification of whether the observational input is consistent with the causal hypothesis. However, the notion of compatibility at play might be more refined than this, for instance, the algorithm might specify a confidence level with which one can reject the given causal hypothesis as an explanation of the observational input.

The inflation DAG technique is a way of mapping a given causal inference problem onto a new causal inference problem where the observational and causal inputs of the latter problem are determined by the observational and causal inputs of the original problem. In particular, under the inflation map, the observational input becomes less informative while the causal hypothesis becomes more fine-grained, in a manner that we shall now make explicit.

III. INFLATION: A TOOL FOR CAUSAL INFERENCE

We now introduce the notion of **an inflation of a causal model**. Let C_G denote a causal model associated to a DAG G . An inflation of this model is another causal model, denoted $C_{G'}$ and associated to a different DAG, G' . We refer to G' as an inflation of G . There are many possible choices of G' for a given G (specified below), hence many possible inflations of a given causal model. We denote the set of these by $\text{Inflations}[G]$. Once an element $G' \in \text{Inflations}[G]$ is specified, however, the parameters of the causal model $C_{G'}$ are fixed by the parameters of the causal model C_G . We denote this function by $C(G') = \text{Inflation}[C(G)]$.

We begin by defining the condition under which a DAG G' is an inflation of a DAG G .

For each node of G , G' may contain one or more copies of that node. If A denotes a node in the DAG G that has copies in G' , then we denote these copies by A_1, \dots, A_k . The variable that indexes the copies is termed the *copy-index*. The necessary condition on G' for being an inflation of G is that the ancestral subgraph in G' of a node A_i is equivalent, under removal of the copy-index, to the ancestral subgraph in G of the node A . When two objects (e.g. nodes, sets of nodes, DAGs, etc. . .) are the same up to copy-indices, then we use \sim to indicate this. For instance, we have $A_i \sim A_j \sim A$. Given this notational convention, we can formalize the condition for G' to be an inflation of G as follows:

$$G' \in \text{Inflations}[G] \quad \text{iff} \quad \forall A_i \in \text{Nodes}[G'] : \text{AnSubDAG}_{G'}[A_i] \sim \text{AnSubDAG}_G[A]. \quad (3) \quad \boxed{\text{eq:definition}} \quad \text{housey2014gdag, woodspekkens, frit22012bell}$$

To illustrate the notion of the inflation of a DAG, we consider the DAG of Fig. 1, which is called the *Triangle scenario* (for obvious reasons) and which has been studied by many authors [7 (Fig. E#8), 5 (Fig. 18b), 6 (Fig. 3), 11 (Fig. 6a), 20 (Fig. 1a), 21 (Fig. 8), 22 (Fig. 1b), 12 (Fig. 4b)] Different inflations of the Triangle scenario are depicted in Figs. 2 to 6.

Next, we turn to specifying how the parameters in the inflation of a causal model are determined from those in the original model.

The condition is simply that for every node A_i in the inflation DAG G' , the manner in which A_i depends causally on its parents within G' must be the same as the manner in which A depends causally on its parents within the original DAG G . Formally,

$$C' \in \text{Inflations}[C] \quad \text{iff} \quad G' \in \text{Inflations}[G] \quad \text{and} \quad \forall A_i \in \text{Nodes}[G'] : P_{C'}(A_i | \text{Pa}_{G'}(A_i)) = P_C(A | \text{Pa}_G(A)), \quad (4) \quad \boxed{\text{eq:function}} \quad \text{leq:definitionDAG}$$

where Eq. (3) guarantees that $A_i \sim A$ and $\text{Pa}_{G'}(A_i) \sim \text{Pa}_G(A)$.

To sum up then, inflation is a mapping to a new causal model wherein each given variable in the original DAG may have counterparts in the inflation DAG and where the causal dependences in the inflation DAG are given by the

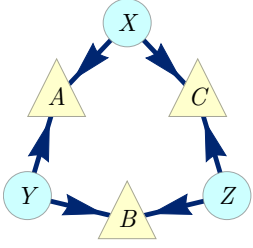


FIG. 1. The causal structure of the Triangle scenario.

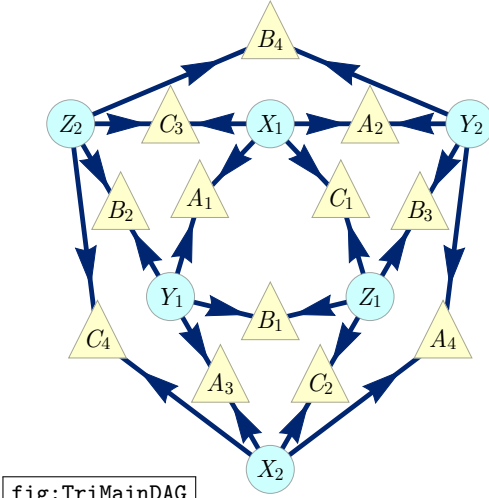


FIG. 2. An inflation DAG of the Triangle scenario where each latent node has been duplicated, resulting in four copies of each observable node.

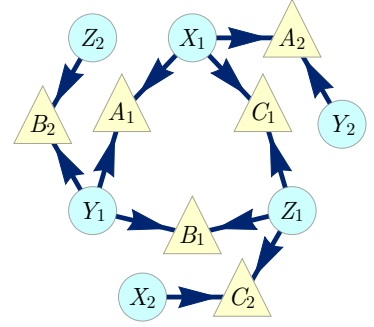


FIG. 3. Another inflation of the Triangle scenario consisting, also notably $\text{AnSubDAG}_{(\text{Fig. 2})}[A_1 A_2 B_1 B_2 C_1 C_2]$.

corresponding causal dependences in the original DAG. Note that the operation of modifying a DAG and equipping the modified version with causal dependences that mirror those of the original also appears in the *do calculus* of [Pearl \[1\]](#). [Pearl 2009 causality](#) Fixme Note: Because causal hypotheses are sets of causal models, any inflation map, by virtue of being a map on causal models, induces a map on causal hypotheses. Specifically,

$$H_{G',S'} \in \text{Inflations}[H_{G,S}] \quad \text{iff} \quad G' \in \text{Inflations}[G] \quad \text{and} \quad H_{G',S'} = \{C' : C' = \text{Inflation}[C], C \in H_{G,S}\}. \quad (5)$$

Note that even if $S = S_{\text{full}}$, so that the original causal hypothesis puts no constraints on the parameter values, one generally has $S' \neq S'_{\text{full}}$, that is, the inflationary image of the full set of parameter values on G is not the full set of parameter values on G' . Rather, S' incorporates a nontrivial restriction on the parameters consistent with G' , namely, that if nodes A_i and A_j on G' are copies of a single node A on G , then the parameter values on G' are constrained to satisfy $P(A_i | \text{Pa}_{G'}(A_i)) = P(A_j | \text{Pa}_{G'}(A_j))$.

To see how inflation is relevant for causal inference, we must explain how the distributions that can be achieved in the inflation model are constrained by those in the original model. In what follows, we assume that $G' \in \text{Inflations}[G]$ and that $C' = \text{Inflation}[C]$.

Note, first of all, that for any sets of nodes $U \in \text{Nodes}[G']$ and $\tilde{U} \in \text{Nodes}[G]$,

$$\text{if } \text{AnSubDAG}_{G'}[U] \sim \text{AnSubDAG}_G[\tilde{U}] \quad \text{then} \quad P_{C'}(U) = P_C(\tilde{U}). \quad (6) \quad \{\text{eq:coinc}\}$$

This follows from the fact that the probability distributions over U and \tilde{U} depend only on their ancestral subgraphs

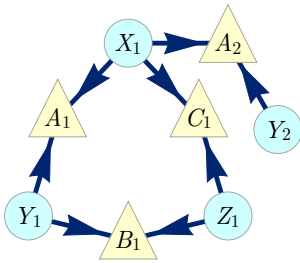


FIG. 4. A simple inflation of the Triangle scenario, also notably $\text{AnSubDAG}_{(\text{Fig. 3})}[A_1 A_2 B_1 C_1]$.

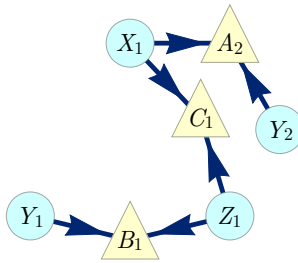


FIG. 5. An even simpler inflation of the Triangle scenario, also notably $\text{AnSubDAG}_{(\text{Fig. 4})}[A_1 A_2 B_1 C_1]$.

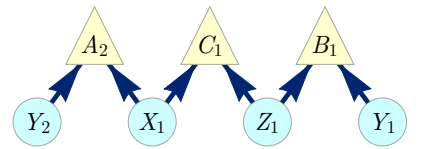


FIG. 6. Another representation of Fig. 5. Despite not containing the original scenario, this is a valid inflation per Eq. (3).

and the parameters defined thereon, which by the definition of inflation are the same for \mathbf{U} and for $\tilde{\mathbf{U}}$.

It is useful to have a name for a set of nodes in the inflation DAG, $\mathbf{U} \subseteq \text{Nodes}[G']$, such that one can find a corresponding set in the original DAG, $\tilde{\mathbf{U}} \subseteq \text{Nodes}[G]$, which has an equivalent ancestral subgraph. We say that such subsets of the nodes of G' are injectable into G and we call them the **injectable sets**,

$$\mathbf{U} \in \text{InjectableSets}[G'] \quad \text{iff} \quad \exists \tilde{\mathbf{U}} \subseteq \text{Nodes}[G] : \text{AnSubDAG}_{G'}[\mathbf{U}] \sim \text{AnSubDAG}_G[\tilde{\mathbf{U}}]. \quad (7) \quad \{\text{eq:defin}$$

In Fig. 3, for example, the set $\{A_1 B_1 C_1\}$ is injectable because its ancestral subgraph is equivalent up to copy-indices to the ancestral subgraph of $\{ABC\}$ in the original DAG (which is just the full DAG), and the set $\{A_2 C_1\}$ is injectable because its ancestral subgraph is equivalent to that of $\{AC\}$ in the original DAG.

Note that it is clear that a set of nodes in the inflation DAG can only be injectable if it contains at most one copy of any node from the original DAG. Similarly, it can only be injectable if its ancestral subgraph also contains at most one copy of any node from the original DAG. Thus, in Fig. 3, $\{A_1 A_2 C_1\}$ is not injectable because it contains two copies of A , and $\{A_2 B_1 C_1\}$ is not injectable because its ancestral subgraph contains two copies of Y .

The sets of nodes of G' that will be of primary interest to us are not the injectable sets per se, but sets satisfying a slightly weaker constraint. To define the latter sorts of sets, which we term *pre-injectable*, we must first introduce some additional terminology.

We refer to a pair of nodes which do not share any common ancestor as being **ancestrally independent**, for which we invent the notation $X \not\sim_{\emptyset} Y$. Generalizing to sets, $\mathbf{U} \not\sim_{\emptyset} \mathbf{V}$ indicates that no node in \mathbf{U} shares a common ancestor with any node in \mathbf{V} ,

$$\mathbf{U} \not\sim_{\emptyset} \mathbf{V} \quad \text{iff} \quad \text{An}(\mathbf{U}) \cap \text{An}(\mathbf{V}) = \emptyset. \quad (8)$$

Ancestral independence is equivalent to d -separation by the empty set [\[1–4\]](#). [pearl2009causality,spirtes2011causation,studeny2005probabil](#)

A set of nodes \mathbf{U} in the inflation DAG G' will be called **pre-injectable** whenever it is a union of injectable sets with disjoint ancestries.

$$\mathbf{U} \in \text{PreInjectableSets}[G'] \quad \text{iff} \quad \exists \{\mathbf{U}_i \in \text{InjectableSets}[G']\} \quad \text{s.t.} \quad \mathbf{U} = \bigcup_i \mathbf{U}_i \quad \text{and} \quad \forall i \neq j : \mathbf{U}_i \not\sim_{\emptyset} \mathbf{U}_j. \quad (9) \quad \{\text{eq:defin}$$

Note that every injectable set is a trivial example of a pre-injectable set.

Because ancestral independence in the DAG implies statistical independence for any probability distribution consistent with the DAG, it follows that if \mathbf{U} is a pre-injectable set and $\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_n$ are the ancestrally independent components of \mathbf{U} , then

$$P_{C'}(\mathbf{U}) = P_{C'}(\mathbf{U}_1) P_{C'}(\mathbf{U}_2) \cdots P_{C'}(\mathbf{U}_n). \quad (10) \quad \{\text{eq:prein}$$

Furthermore, because each injectable set of variables \mathbf{U}_i satisfies Eq. (6), it follows that joint distributions on pre-injectable sets in the inflation model can be expressed in terms of distributions defined on the original causal model, [eq:coincidingdistrodef](#)

$$P_{C'}(\mathbf{U}) = P_C(\tilde{\mathbf{U}}_1) P_C(\tilde{\mathbf{U}}_2) \cdots P_C(\tilde{\mathbf{U}}_n). \quad (11) \quad \{\text{eq:prein}$$

This latter property implies that the probability distribution over any pre-injectable set in the inflation model is fully determined by the parameters in the original causal model. Indeed, this relation is what motivates us to consider the pre-injectable sets.

Consider a causal inference problem where the observational input is a probability distribution over a set \mathbf{O} of observed variables, denoted $P(\mathbf{O})$ and the causal hypothesis is $H_{G,S}$ where $\text{ObservedNodes}[G] = \mathbf{O}$. Suppose that one seeks only to determine whether the causal hypothesis is consistent with the observational input. This holds if and only if

$$\exists C \in H_{G,S} : P_C(\text{ObservedNodes}[G]) = P(\mathbf{O}). \quad (12)$$

Every inflation $G \rightarrow G'$ defines a new causal inference problem for which the decision regarding consistency is the same as for the original problem.

The causal hypothesis of the new causal inference problem is simply the image under a $G \rightarrow G'$ inflation of the causal hypothesis of the original causal inference problem, and one seeks to evaluate its consistency with observational data that is determined by the observational data of the original in the following way.

Let \mathbf{O} denote the image of $\tilde{\mathbf{O}}$ under the $G \rightarrow G'$ inflation, and consider the subsets of \mathbf{O} that are pre-injectable relative to the $G \rightarrow G'$ inflation, denoted $\text{PreInjectableSets}[\mathbf{O}]$. Recall that for every set of nodes $\mathbf{U} \in \text{PreInjectableSets}[\mathbf{O}]$

there is a partition $\mathbf{U} = \bigcup_i \mathbf{U}_i$ where the \mathbf{U}_i are subsets of \mathbf{O} that are injectable relative to the $G \rightarrow G'$ inflation and that are ancestrally independent. Recall also that, by the definition of a $G \rightarrow G'$ inflation, for any set of nodes $\mathbf{U}_i \subseteq \mathbf{O}$ that is injectable, one can find a set of nodes $\tilde{\mathbf{U}}_i \subseteq \tilde{\mathbf{O}}$ such that $\forall i : \mathbf{U}_i \sim \tilde{\mathbf{U}}_i$. The observational data of the new causal inference problem is that $\forall \mathbf{U} \in \text{PreInjectableSets}[\mathbf{O}] : P(\mathbf{U}) = P(\tilde{\mathbf{U}}_1)P(\tilde{\mathbf{U}}_2) \dots P(\tilde{\mathbf{U}}_n)$, where $P(\tilde{\mathbf{U}}_i)$ is the marginal of $P(\tilde{\mathbf{O}})$ on $\tilde{\mathbf{U}}_i$.

Therefore the observational input to the new causal inference problem is not the distribution on \mathbf{O} , but rather a specification of the marginals of this distribution on each of the pre-injectable sets of \mathbf{O} under the $G \rightarrow G'$ inflation.

Summarizing, we have:

Lemma 1. *The causal hypothesis $H_{G,S}$ is consistent with the observational data $P(\tilde{\mathbf{O}})$ if and only if the image of this causal hypothesis under a $G \rightarrow G'$ inflation, denoted $H_{G',S'}$, is consistent with the observational data that $\forall \mathbf{U} \in \text{PreInjectableSets}[\mathbf{O}] : P(\mathbf{U}) = P(\tilde{\mathbf{U}}_1)P(\tilde{\mathbf{U}}_2) \dots P(\tilde{\mathbf{U}}_n)$.*

It follows that any consistency conditions that one can derive for the new causal inference problem immediately yield consistency conditions for the original causal inference problem. Indeed, any standard tool of causal inference can be applied to the new problem and conditions derived therefrom provide novel conditions for the original problem. Any causal inference tool, therefore, can potentially have its power augmented by combining it with inflation.

Example 1: Perfect correlation cannot arise from the Triangle scenario.

Consider the following causal inference problem. The observational data is a joint distribution over three binary-outcome variables, $P(ABC)$, where the marginal on each variable is uniform and the three are perfectly correlated,

$$p(abc) = \frac{[000] + [111]}{2} = \begin{cases} \frac{1}{2} & \text{if } a = b = c, \\ 0 & \text{otherwise.} \end{cases} \quad (13) \quad \boxed{\text{eq:ghzdist}}$$

The causal hypothesis is the set of causal models associated to the DAG of the triangle scenario, depicted in Fig. 1. The problem is to decide whether the observational data is consistent with this causal hypothesis.

To solve this problem, we consider the causal inference problem induced by the particular inflation of the triangle scenario depicted in Fig. 6.

The sets $\{A_2C_1\}$ and $\{B_1C_1\}$ are injectable (hence trivially pre-injectable). The set $\{A_2B_1\}$ is also pre-injectable because the singleton sets $\{A_2\}$ and $\{B_1\}$ are each injectable and ancestrally independent. Therefore, the observational data in our new causal inference problem includes

$$P(A_2C_1) = P(AC) \quad (14) \quad \boxed{\text{j1}}$$

$$P(B_1C_1) = P(BC) \quad (15) \quad \boxed{\text{j2}}$$

$$P(A_2B_1) = P(A)P(B). \quad (16) \quad \boxed{\text{j3}}$$

The question is whether such constraints are consistent with the causal hypothesis that the DAG is that of Fig. 1, where the parameters supplementing the DAG are arbitrary, except for the constraint that the latent variables Y_1 and Y_2 are identically distributed.

It is not difficult to see that the answer to the question is negative. And this verdict can be rendered without even appealing to the form of the inflation DAG (Note, however, that the inflation DAG has played a role in defining the new causal inference problem insofar as it has dictated the form that the observational data should take). It suffices to make use of results concerning the marginals problem. From Eq. (13) one deduces that the marginal $P(AC)$ described perfect correlation between A and C and similarly for the marginal $P(BC)$. But then, by Eqs. (14) and (15), one deduces that there is perfect correlation between A_2 and C_1 and perfect correlation between B_1 and C_1 . Meanwhile, Eq. (16) dictates that A_2 and B_1 should be uncorrelated. But there is no joint distribution $P(A_2B_1C_1)$ having such marginals: the only joint distribution that exhibits perfect correlation between A_2 and C_1 and between B_1 and C_1 also exhibits perfect correlation between A_2 and B_1 .

We have therefore certified that perfect correlations among A , B and C is not consistent with the Triangle causal structure, recovering the seminal result of Steudel and Ay [22].

Example 2: The W-type distribution cannot arise from the Triangle scenario

Consider another causal inference problem concerning the triangle scenario, namely, that of determining whether the triangle DAG can explain a joint distribution $P_W(ABC)$ of the form

$$p_W(abc) = \frac{[100] + [010] + [001]}{3} = \begin{cases} \frac{1}{3} & \text{if } a + b + c = 1, \\ 0 & \text{otherwise.} \end{cases} \quad (17) \quad \boxed{\text{eq:wdist}}$$

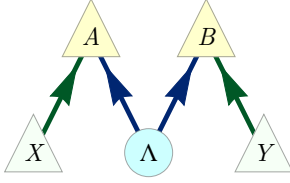


FIG. 7. The causal structure of the bipartite Bell scenario. The local outcomes of Alice's and Bob's experimental probing is assumed to be a function of some latent common cause, in addition to their independent local experimental settings.

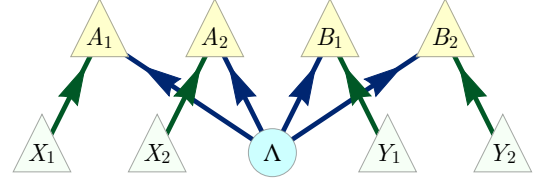


FIG. 8. An inflation DAG of the bipartite Bell scenario, where both local settings variables have been duplicated.

fig:BellD

fig:NewBellDAG1

The correlations described by this distribution are known as PR-box correlations after Popescu and Rohrlich, and they are well-known to be inconsistent with the the Bell scenario [37, 38]. Here we prove this inconsistency using the inflation technique.

We use the inflation of the Bell DAG shown in Fig. 8.

We begin by recognizing that $\{A_1 B_1 X_1 Y_1\}$, $\{A_2 B_1 X_2 Y_1\}$, $\{A_1 B_2 X_1 Y_2\}$, and $\{A_2 B_2 X_2 Y_2\}$ are all injectable sets, and that $\{X_1 X_2 Y_1 Y_2\}$ is a pre-injectable set because the singleton sets $\{X_1\}$, $\{X_2\}$, $\{Y_1\}$, and $\{Y_2\}$ are all injectable and ancestrally independent of one another. It follows that

$$P(A_1 B_1 X_1 Y_1) = P(ABXY) \quad (29) \quad \{\text{PR1}\}$$

$$P(A_2 B_1 X_2 Y_1) = P(ABXY) \quad (30) \quad \{\text{PR2}\}$$

$$P(A_1 B_2 X_1 Y_2) = P(ABXY) \quad (31) \quad \{\text{PR3}\}$$

$$P(A_2 B_2 X_2 Y_2) = P(ABXY) \quad (32) \quad \{\text{PR4}\}$$

$$P(X_1 X_2 Y_1 Y_2) = P(X)P(X)P(Y)P(Y). \quad (33) \quad \{\text{PR5}\}$$

These imply that

$$P(A_1 B_1 | X_1 Y_1) = P(AB | XY) \quad (34) \quad \{\text{PR1b}\}$$

$$P(A_2 B_1 | X_2 Y_1) = P(AB | XY) \quad (35) \quad \{\text{PR2b}\}$$

$$P(A_1 B_2 | X_1 Y_2) = P(AB | XY) \quad (36) \quad \{\text{PR3b}\}$$

$$P(A_2 B_2 | X_2 Y_2) = P(AB | XY). \quad (37) \quad \{\text{PR4b}\}$$

Given the assumption that the distributions $P(X)$ and $P(Y)$ are full support, it follows from Eq. (33) that sometimes $\{X_1, X_2, Y_1, Y_2\} = \{0, 1, 0, 1\}$. For these values, P_{PR} and Eq. (34) specifies perfect correlation between A_1 and B_1 . Similarly, for these values we expect perfect correlation between A_1 and B_2 , perfect correlation between A_2 and B_1 , and perfect anticorrelation between A_2 and B_2 . Those four requirements are not mutually compatible, however: since perfect correlation is transitive, the first three properties entail perfect correlation between A_2 and B_2 .

The mathematical structure of this proof parallels that of standard proofs of the inconsistency of PR-box correlations with the Bell structure. Standard proofs focus on a set of variables $A_0, A_1, B_0, B_1, C_0, C_1$ where A_0 is the value of A when $X = 0$, and similarly for the others. The fact that there must be a joint distribution over these variables can be inferred from the structure of the Bell DAG and the fact that one can assume without loss of generality that the causal dependences are deterministic (a result known as Fine's theorem [citation]). It is then sufficient to note that the marginals given by the PR-box correlations do not arise from any joint distribution. Nonetheless, our proof is conceptually distinct insofar as the variables to which we apply the marginals problem are not conditioned on a setting value. And we do not require Fine's theorem. [Can we do something that couldn't be done in the standard approach?]

IV. A PROCEDURE FOR DERIVING POLYNOMIAL INEQUALITIES

Our final goal in deriving polynomial inequalities is to derive constraints on the distribution of the observable variables in the original DAG. We accomplish this by proxy, namely by deriving constraints on the distributions over the variables in the pre-injectable sets of the inflation DAG. Just as distributions over the pre-injectable sets translate into distributions pertaining to the original DAG variables per Eq. (6), so do *constraints* on the distributions

sec:ineqs

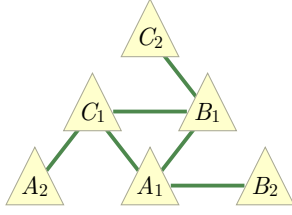


FIG. 9. The auxiliary injection graph corresponding to the inflation DAG in Fig. 3, wherein a pair of nodes are adjacent iff they are pairwise injectable.

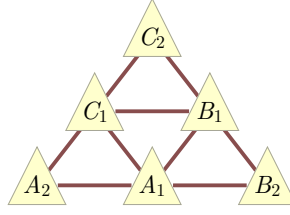


FIG. 10. An auxiliary ancestral dependence graph corresponding to the inflation DAG in Fig. 3, wherein a pair of nodes are adjacent iff they share a common ancestor.

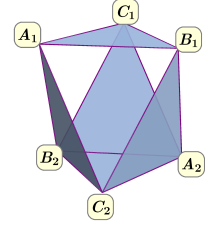


FIG. 11. The simplicial complex... The 5 faces in this figure correspond to the pre-injectable sets.

of pre-injectable sets translate into constraints on the original DAG variables.

We are aware of a variety of methods for identifying interesting constraints on the distributions over the variables in the pre-injectable sets. But here, we focus on the very simplest and weakest such constraint: the nonnegativity of probabilities together with the factorization Eq. (11). In fact, this is all that we have used in the previous section as well.

Our method is based on deriving inequalities for the so-called **marginal problem**. This relies on linear quantifier elimination, which is computationally feasible in at least some cases. A nonlinear method for deriving even tighter constraints is discussed in appendix Appendix A. The simpler method discussed here consists of two steps: 1) Identify the pre-injectable sets, and 2) Solve the marginal problem with respect to the pre-injectable sets.

Step 1: Identifying the Pre-Injectable Sets

To identify the pre-injectable sets, we first identify the injectable sets. To this end, it is useful to construct an auxiliary graph from the inflation DAG. Let the nodes of these auxiliary graphs be the observable nodes in the inflation DAG. The **injection graph**, then, is the undirected graph in which a pair of nodes A_i and B_j are adjacent if $An(A_i B_j)$ is irredundant. The injectable sets are then precisely the cliques⁵ in this graph, per ??.

Determining the pre-injectable sets from there can be done via constructing another graph that we call the **independence graph**. Its nodes are the injectable sets, and we connect two of these by an edge if their ancestral subgraphs are disjoint. Then by definition, the pre-injectable sets can be obtained as the cliques in this graph. Taking the union of all the injectable sets in such a clique results in a pre-injectable set. Since it is sufficient to only consider the maximal pre-injectable sets, one can eliminate all those pre-injectable sets that are contained in other ones, as a final step.

Let us also define the **ancestral dependence graph**, in which two nodes are adjacent if they share a common ancestor, and its complement the **ancestral independence graph**, in the ancestrally independent nodes are adjacent. To ascertain the factorization of a node set U into ancestrally-independent partitions one considers the subgraph on U of the ancestral dependence graph: the ancestrally-independent partitions are identically the distinct connected components of that subgraph. By examining the injection graph and the ancestral dependence graph, therefore, one is able to quickly determine all injectable sets and all ancestral independence relations.

Applying these prescriptions to the inflation DAG in Fig. 3 identifies the following ancestral independences maximal injectable and maximal pre-injectable sets as follows:

$$\begin{array}{cccccc}
 A_2 \not\sim B_1 & & \{A_1 B_1\} & & \{A_1 B_1 C_1\} & \\
 A_2 \not\sim C_2 & & \{B_1 C_1\} & & \{A_1 B_2 C_2\} & \\
 B_2 \not\sim A_2 & & \{A_1 C_1\} & & \{A_2 B_1 C_2\} & \\
 B_2 \not\sim C_1 & & \{A_2 C_1\} & & \{A_2 B_2 C_1\} & \\
 C_2 \not\sim A_1 & & \{B_2 A_1\} & & \{A_2 B_2 C_2\} & \\
 C_2 \not\sim B_2 & & \{C_2 B_1\} & & & \\
 \underbrace{\hspace{10em}}_{\text{pairwise ancestral independences}} & \underbrace{\{A_2\} \not\sim \{B_1 B_2 C_2\}, \{B_2\} \not\sim \{A_2 C_1 C_2\}, \{C_2\} \not\sim \{A_1 A_2 B_2\}, \{A_2\} \not\sim \{B_2\} \not\sim \{C_2\}}_{\text{maximal ancestral independences}} & \underbrace{\hspace{10em}}_{\text{pairwise injectable sets}} & \underbrace{\{A_1 B_2\}, \{B_1 C_2\}, \{A_2 C_1\}, \{A_1 B_1 C_1\}}_{\text{maximal injectable sets}} & \underbrace{\hspace{10em}}_{\text{maximal pre-injectable sets}} & (38) \quad \text{fig:basic}
 \end{array}$$

⁵ A clique is a set of nodes such that every single node is connected to every other.

such that the distributions on the pre-injectable sets relate to the original DAG distribution via

$$\begin{aligned}
P(A_1 B_1 C_1) &= P(ABC) \\
P(A_1 B_2 C_2) &= P(C)P(AB) \\
P(A_2 B_1 C_2) &= P(A)P(BC) \\
P(A_2 B_2 C_1) &= P(B)P(AC) \\
P(A_2 B_2 C_2) &= P(A)P(B)P(C)
\end{aligned} \tag{39}$$

Having identified the pre-injectable sets (and how to use them), we next turn to deriving constraints on the distributions over the pre-injectable sets.

Step 2: Constraining the Distribution over Pre-Injectable Sets via the Marginal Problem

The most trivial constraint possible is the *existence of a joint probability distribution* over all the observable variables in the inflation DAG. Each of the five three-variable distributions in Eq. (39) is a different marginal distribution of the six-variable joint distribution $P(A_1 A_2 B_1 B_2 C_1 C_2)$. Solving the marginal problem means finding inequalities on the marginal distributions such that the inequalities will be satisfied only if there exists some joint distribution from which the distributions on the marginal sets can be recovered through marginalization. The marginal problem comes up in a variety of applications, and has been studied extensively; see [10] for further references. Tobias, say better? And add citations please! T: Ok, I'll get back to this on my second iteration

Comment about how solving marginals problem still gives polynomial inequalities.

From now on, one can apply *any* method for dealing with the marginal problem. For example if one is given a particular distribution P over the original scenario variables, then one can use Eq. (11) to compute the marginal distributions over the pre-injectable sets in an inflation DAG. Efficient linear programming can be used to check whether there exists a joint distribution reproducing these marginals [39, 40]. If such a joint distribution does not exist, then the given original-scenario distribution is witnessed as inconsistent with the original causal structure. If a joint distribution *does* exist then the problem remains open, and one can try again using a different inflation.

In order to derive inequalities that hold for *all* distributions that are feasible on the original DAG, say for a given number of outcomes per variable, we can formally solve the marginal problem via linear quantifier elimination, as we review in the following. This consists of eliminating unknowns from a set of linear inequalities and equalities.

The linear inequalities correspond to the nonnegativity of the joint distribution. Formally, the probability of any possible assignment of outcomes to the observable variables is constrained to be nonnegative. For the six observable variables in Fig. 3, for example, the linear inequalities are given by $0 \leq P(A_1 A_2 B_1 B_2 C_1 C_2)$. Note that a single inequality (or equality) pertaining to a probability *distribution* is actually shorthand for a whole set of inequalities (or equalities) pertaining to the probabilities of individual events. Taking the observable variables to be binary, for example, would mean that $0 \leq P(A_1 A_2 B_1 B_2 C_1 C_2)$ would be shorthand for 64 distinct nonnegativity inequalities, namely

$$\begin{aligned}
0 &\leq p(a_1 a_2 b_1 b_2 c_1 c_2), & 0 &\leq p(\bar{a}_1 a_2 b_1 b_2 c_1 c_2), & 0 &\leq p(a_1 \bar{a}_2 b_1 b_2 c_1 c_2), & \dots, \\
0 &\leq p(\bar{a}_1 \bar{a}_2 b_1 b_2 c_1 c_2), & 0 &\leq p(\bar{a}_1 a_2 \bar{b}_1 b_2 c_1 c_2), & \dots
\end{aligned} \tag{40}$$

etc. For a marginals problem based on the joint existence of n observable variables, each ranging over r possible outcomes, one would initialize the problem with r^n linear nonnegativity inequalities. Unlike probabilities pertaining to injectable sets, these joint probabilities are *not* fully specified by probabilities which are observed in the original scenario. We coin the term **gedankenprobability** to denote a probability pertaining to a not-pre-injectable set of inflation-DAG variables. The gedankenprobabilities evoke thought experiments, because knowing the original-scenario causal model determines the gedankenprobabilities.

The linear equations in the marginal problem are those equations which express each of the marginal probabilities as a sum over various different gedankenprobabilities, namely marginalization. Again using distribution shorthand, the

five marginal distributions in Eq. (39) would correspond to the equations

$$\begin{aligned}
 P(A_1 B_1 C_1) &= \sum_{A_2 B_2 C_2} P(A_1 A_2 B_1 B_2 C_1 C_2), \\
 P(A_1 B_2 C_2) &= \sum_{A_2 B_1 C_1} P(A_1 A_2 B_1 B_2 C_1 C_2), \\
 P(A_2 B_1 C_2) &= \sum_{A_1 B_2 C_1} P(A_1 A_2 B_1 B_2 C_1 C_2), \\
 P(A_2 B_2 C_1) &= \sum_{A_1 B_1 C_2} P(A_1 A_2 B_1 B_2 C_1 C_2), \\
 P(A_2 B_2 C_2) &= \sum_{A_1 B_1 C_1} P(A_1 A_2 B_1 B_2 C_1 C_2).
 \end{aligned} \tag{41}$$

To be clear, taking the observable variables to be binary would mean that each equality in Eqs. (41) is shorthand for 8 distinct marginalization equations. For example the equality $P(A_1 B_1 C_1) = \sum_{A_2 B_2 C_2} P(A_1 A_2 B_1 B_2 C_1 C_2)$ is shorthand for 8 equations of the form

$$p(A_1=a, B_1=b, C_2=c) = \sum_{a' b' c'} p(a_1 a' b_1 b' c_1 c'_2) \tag{42}$$

for each of the 8 possible values of the tuple $\{abc\}$.

Solving the marginal problem means eliminating all the gedankenprobabilities such as $p(a_1 a' b_1 b' c_1 c'_2)$ from the system of inequalities and equalities. Practically, this is accomplished by linear quantifier elimination. As an equivalent description of the same problem, but could likewise consider the problem of determining the facets of the marginal polytope, the vertices of which are known to be given by the deterministic points (compare Fine’s theorem [41] in the Bell scenario case).

Geometrically, linear quantifier elimination is equivalent to projecting a high-dimensional polytope in halfspace representation (inequalities and equalities) into a lower-dimensional quotient space.

Polytope projection is a well-understood problem in computational optimization, and a surprising variety of algorithms are available for the task [42–44]. The oldest-known method for polytope projection, i.e. linear quantifier elimination, is an algorithm known as Fourier-Motzkin (FM) elimination [45, 46] although Fourier-Cernikov elimination variant [47, 48], as well as Block Elimination and Vertex Enumeration [49], are also fairly popular. More advanced polytope projection algorithms, such as Equality Set Projection (ESP) [42–44] and Parametric Linear Programming, have also recently become available [42–44].

Linear quantifier elimination routines are available in many software tools⁶. We have found custom-coding an linear elimination routine in *Mathematica*TM to be most efficient, see Appendix B for further detail.

Linear quantifier elimination is already widely used in causal inference to derive entropic inequalities [10–12]. 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 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. It is therefore sometimes useful to consider relaxations of the marginals problem. The *full* marginals problem is to find inequalities on the marginal distributions such that the inequalities are satisfied *if and only if* the given marginal distributions can be extended. It is much easier to generate necessary-but-insufficient inequalities, i.e. satisfied by all compatible marginal distributions but such that no-violation does not certify compatibility. We have identified a technique for rapidly generating such quantifier-free inequalities by restricting the search to inequalities of a very particular form. We found this alternative technique — trading generality for speed — to be extraordinarily practical. The type of inequalities that we consider are given by a certain class of tautologies in classical propositional logic, see Sec. VII for further details.

As far as we can tell, our inequalities are not related to the nonlinear infeasibility criteria which have been derived specifically to constrain classical networks [51–53] nor to the nonlinear inequalities which account for interventions to a given causal structure [36, 54].

⁶ For example *MATLAB*TM’s `MPT2/MPT3`, *Mazima*’s `fourier.elim`, *lrs*’s `fourier`, or *Maple*TM’s (v17+) `LinearSolve` and `Projection`. The efficiency of most of these software tools, however, drops off markedly when the dimension of the final projection is much smaller than the initial space of the inequalities. FM elimination aided by Cernikov rules [47, 48] is implemented in *qskeleton* [50]. ESP [42–44] is supported by `MPT2` but not `MPT3`, and by the (undocumented) option of `projection` in the *polytope* (v0.1.1 2015-10-26) python module.

V. EXAMPLES OF POLYNOMIAL INEQUALITIES FOR THE TRIANGLE SCENARIO

T: why not just refer to the list of ineqs here, state that we've done a successful quantifier elimination to demonstrate its practicality, and discuss one or two of the most interesting ineqs?

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

The nontrivial polynomial inequality

$$p(a) + p(b) + p(c) \leq 1 + p(a)p(b) + p(ac) + p(bc) \quad (43) \quad \text{\texttt{\{eq:tritr}}}$$

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

$$\begin{aligned} 0 &\leq 1 - p(a_2) - p(b_1) - p(c_1) + p(a_2b_1) + p(a_2c_1) + p(b_1c_1) - p(a_2b_1c_1) \quad [= p(\bar{a}_2\bar{b}_1\bar{c}_1)] \\ 0 &\leq p(a_2b_1c_1) \end{aligned} \quad (44) \quad \text{\texttt{\{eq:trisi}}$$

subject to the following transformations

$$\underbrace{p(a_2b_1) \rightarrow p(a_2)p(b_1)}_{\text{Factorization relation, per Eqs. (A-4).}} \quad \text{and} \quad \underbrace{\begin{aligned} &p(a_2) \rightarrow p(a), p(b_1) \rightarrow p(b), p(c_1) \rightarrow p(c), \\ &p(a_2c_1) \rightarrow p(ac), p(b_1c_1) \rightarrow p(bc) \end{aligned}}_{\text{Mapping relations, per Eqs. (A-6)}}. \quad (45)$$

Indeed, this example has been chosen because Eq. (43) can be derived directly from the small ancestral subgraph of $\{A_2B_1C_1\}$, namely Fig. 6.

It may be constructive to rewrite this inequality in correlator form using outcomes in $\{-1, +1\}$ as

$$\langle AC \rangle - \langle BC \rangle \leq 1 - \langle A \rangle \langle B \rangle,$$

which indicates that like ??, it is also a sort of monogamy inequality: it is impossible for A and C to be strongly correlated while B and C are strongly anticorrelated.

A consequence of Eq. (43) is that the perfect correlation distribution per Eq. (13) is found to be unrealizable from the Triangle scenario. This conclusions follows by considering the special case of Eq. (43) where $a \rightarrow 1, b \rightarrow 1, c \rightarrow 0$.

Slightly more involved but otherwise analogous considerations give rise to the inequality

$$p(a_2b_2c_2) \leq p(\bar{a}_1\bar{b}_1\bar{c}_1) + p(a_1b_2c_2) + p(a_2b_1c_2) + p(a_2b_2c_1) \quad (46) \quad \text{\texttt{\{eq:trifa}}$$

which in turn yields

$$p(a)p(b)p(c) \leq p(\bar{a}\bar{b}\bar{c}) + p(c)p(ab) + p(b)p(ac) + p(a)p(bc) \quad (47) \quad \text{\texttt{\{eq:Fritz}}$$

per

$$\underbrace{\begin{aligned} &p(a_1b_2c_2) \rightarrow p(c_2)p(a_1b_2) \\ &p(a_2b_1c_2) \rightarrow p(a_2)p(b_1c_2) \\ &p(a_2b_2c_1) \rightarrow p(b_2)p(a_2c_1) \\ &p(a_2b_2c_2) \rightarrow p(a_2)p(b_2)p(c_2) \end{aligned}}_{\text{Factorization relations}} \quad \text{and} \quad \underbrace{\begin{aligned} &p(a_1b_2) \rightarrow p(ab) \\ &p(a_2c_1) \rightarrow p(ac) \\ &p(b_1c_2) \rightarrow p(bc) \\ &p(\bar{a}_1\bar{b}_1\bar{c}_1) \rightarrow p(\bar{a}\bar{b}\bar{c}) \end{aligned}}_{\text{Nontrivial mapping relations}}. \quad (48)$$

A consequence of Eq. (47) is that the W-type distribution per Eq. (17) is found to be unrealizable from the Triangle scenario, by considering the special case of Eq. (47) where $a \rightarrow 1, b \rightarrow 1, c \rightarrow 1$. Eq. (47) requires the use of a broadcasting inflation, and therefore does not hold in the context of general probability theories.

VI. INEQUALITIES FOR THE BIPARTITE BELL SCENARIO

To further illustrate our inflation-DAG approach, we demonstrate how to recover all Bell inequalities [\[13, 16, 34\]](#) via [Brunner2013Bell, bell19661hv](#) our method. To keep things simple we only discuss the case of a bipartite Bell scenario with two possible “settings”

here, but the cases of more settings and/or more parties are totally analogous.

Consider the causal structure associated to the Bell/CHSH [13, 16, 33, 34] experiment [7 (Fig. E#2), 5 (Fig. 19), 11 (Fig. 1), 8 (Fig. 1), 35 (Fig. 2b), 36 (Fig. 2)], depicted here in Fig. 7. The observable variables are A, B, X, Y , and Λ is the latent common cause of A and B .

In the Bell scenario DAG, one usually works with the conditional distribution $P(AB|XY)$, which is an array of distributions over A and B indexed by the possible values of X and Y , instead of with the original distribution $P(ABXY)$. The maximal pre-injectable sets then are

$$\begin{aligned} &\{A_1 B_1, X_1 X_2 Y_1 Y_2\} \\ &\{A_1 B_2, X_1 X_2 Y_2 Y_2\} \\ &\{A_2 B_1, X_1 X_2 Y_1 Y_2\} \\ &\{A_2 B_2, X_1 X_2 Y_2 Y_2\}, \end{aligned}$$

where we have put commas in order to clarify that every maximal pre-injectable set contains *all* “settings” variables. Using These pre-injectables result in the given marginal distributions to take on the form

$$\begin{aligned} P_{A_1 B_1 X_1 X_2 Y_1 Y_2}(abx_1x_2y_1y_2) &= P_{ABXY}(abx_1y_1)P_X(x_2)P_Y(y_2), \\ P_{A_1 B_2 X_1 X_2 Y_1 Y_2}(abx_1x_2y_1y_2) &= P_{ABXY}(abx_1y_2)P_X(x_2)P_Y(y_1), \\ P_{A_2 B_1 X_1 X_2 Y_1 Y_2}(abx_1x_2y_1y_2) &= P_{ABXY}(abx_2y_1)P_X(x_1)P_Y(y_2), \\ P_{A_2 B_2 X_1 X_2 Y_1 Y_2}(abx_1x_2y_1y_2) &= P_{ABXY}(abx_2y_2)P_X(x_1)P_Y(y_1), \\ P_{X_1 X_2 Y_1 Y_2}(x_1x_2y_1y_2) &= P_X(x_1)P_X(x_2)P_Y(y_1)P_Y(y_2). \end{aligned}$$

Although the last equations follows from each of the others together with the Markov condition $P_{XY} = P_X P_Y$ for the original distribution, it nevertheless turns out to be useful to write down explicitly: putting $x_1 = y_1 = 0$ and $x_2 = y_2 = 1$ and dividing the first equation by the last one results in

$$P_{ABXY}(ab|00) = \sum_{a',b'} P_{A_1 A_2 B_1 B_2 X_1 X_2 Y_1 Y_2}(aa'bb'|0101),$$

and similarly also $P_{ABXY}(ab|01)$, $P_{ABXY}(ab|10)$ and $P_{ABXY}(ab|11)$ can be written as marginals of a conditional distribution. By Fine’s Theorem [41], this implies the existence of a hidden-variable model. Conversely, if a hidden-variable model exists, then the existence of the inflation model implies the existence of a solution to the marginal problem. In conclusion, we therefore find in the case of the inflation DAG Fig. 8, our method provides a necessary and sufficient condition for feasibility with the Bell causal structure. In particular, the marginal problem on this inflation DAG reproduces *all* Bell inequalities.

VII. GEDANKENPROBABILITIES AND THE QUANTUM NO-BROADCASTING THEOREM

It is worth emphasizing that broadcasting gedankenprobabilities are strictly classical constructs. If the latent node in the Bell scenario in Fig. 7 is allowed to be a quantum resource $\mathcal{H}^{d_A \otimes d_B}$, for example, then broadcasting gedankendistributions such as $P(A|x, A|\bar{x}, \dots)$ or $P(A_1, A_2, \dots)$ are **physically prohibited** if the quantum state is suitably entangled.

More precisely, quantum states are governed by a no-broadcasting theorem [23, 24]: 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 measurement setting. As a consequence of the no-broadcasting theorem, in the inflation DAG picture a quantum state which was initially available to a single party cannot be distributed both to Alice-copy-#1 and Alice-copy-#2 in the way a classical hidden variable could be. More generally, there is an analogous no-broadcasting theorem in the regime of epistemically-restricted general probabilistic theories (GPTs) [24, 55–57], so that this impossibility also holds in many theories other than quantum theory.

This means that considerations on inflation DAGs cannot be used to derive quantum causal infeasibility criteria whenever a gedankenprobability presupposes the ability to broadcast a latent node’s system. Broadcasting and non-broadcasting sets of variables are distinguished per ??.

Not every inflation requires broadcasting, however, and hence not every gedankenprobability is physically prohibited by quantum theory. Fig. 6 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. (43), which was derived from

Fig. 6, is therefore a causal infeasibility criterion which holds for the Triangle scenario even when the latent nodes are allowed to be quantum states. This confirms our numerical computations, which indicated that (13) does not have any quantum violations. The same is true for monogamy of correlations, per ???. Since the GHZ-type distribution Eq. (13) violates both of these inequalities, it is forbidden even per the relaxed GPT Triangle scenario, as was pointed out earlier.

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 broadcasting inflation DAG can be characterized in terms of the logical broadcasting maps of Coecke and Spekkens [58]. Note that $p(a_1, a_2)$ and other broadcasting-implicit gedankenprobabilities can be *negative* pursuant to a logical broadcasting map, and hence Step 1 in ??? would need to be modified.

An analysis along these lines has already been carried out successfully by Chaves *et al.* [20] in the derivation of entropic inequalities for non-classical causal structures. Although Chaves *et al.* [20] do not invoke inflation DAGs, they do employ conditional structure, which therefore gives rise to broadcasting sets. Chaves *et al.* [20] take pains to avoid including broadcasting gedankenentropies in any of their initial entropic inequalities, precisely as we would want to do in constructing our initial probability inequalities. Unlike entropic inequalities, the derivation of probability inequalities has not yet been achieved for non-classical causal structures other than Bell scenarios.

Our current inflation DAG method can be employed to derive causal infeasibility criteria for general causal structures, thus generalizing Bell inequalities somewhat. From a quantum foundations perspective, however, generalizing Tsirelson inequalities [16, 59]—the ultimate constraints on what quantum theory makes possible—is even more desirable. Deriving additional quantum causal infeasibility criteria for general causal structure is therefore a priority for future research. **T: move this to conclusions?**

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 quickly happens that even the 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

$$\text{True} \iff \text{Or}[A=a, A \neq a] . \quad (49)$$

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-hand-side while still yielding a valid tautology. Thus

$$\text{And}[\mathbf{a}, \mathbf{b}] \implies \text{Or} \left[\begin{array}{l} \text{And}[\mathbf{a}, \mathbf{c}] , \\ \text{And}[\mathbf{b}, \bar{\mathbf{c}}] \end{array} \right] \quad (50) \quad \{\text{eq:TSEMd}\}$$

is an example of a tautology of the *supplemented* excluded middle (TSEM). 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. (50) is

$$p(ab) \leq p(ac) + p(b\bar{c}) . \quad (51)$$

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

example the TSEM inequality

$$p(a_2b_1) \leq p(a_2b_1) + p(\bar{b}_1c_1) \quad (52)$$

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

$$p(a)p(c) \leq p(ab) + p(\bar{b}c) \quad (53)$$

pursuant to Fig. 6.

In fact almost every causal infeasibility criterion in this paper is a TSEM inequality! The Triangle scenario inequality Eq. (46) corresponds to the logical TSEM

$$\text{And}[\mathbf{a}_2, \mathbf{b}_2, \mathbf{c}_2] \implies \text{Or} \left[\begin{array}{l} \text{And}[\bar{a}_1, \bar{b}_1, \bar{c}_1], \\ \text{And}[a_1, \mathbf{b}_2, \mathbf{c}_2], \\ \text{And}[\mathbf{a}_2, b_1, \mathbf{c}_2], \\ \text{And}[\mathbf{a}_2, \mathbf{b}_2, c_1] \end{array} \right] \quad (54)$$

and then Bell scenario inequality in ?? follows from the logical TSEM

$$\text{And}[\mathbf{a}_1, \mathbf{b}_1, \mathbf{x}_1, \bar{\mathbf{x}}_2, \mathbf{y}_1, \bar{\mathbf{y}}_2] \implies \text{Or} \left[\begin{array}{l} \text{And}[\bar{a}_2, \mathbf{b}_1, \mathbf{x}_1, \bar{\mathbf{x}}_2, \mathbf{y}_1, \bar{\mathbf{y}}_2] \\ \text{And}[\mathbf{a}_1, \bar{b}_2, \mathbf{x}_1, \bar{\mathbf{x}}_2, \mathbf{y}_1, \bar{\mathbf{y}}_2] \\ \text{And}[a_2, b_2, \mathbf{x}_1, \bar{\mathbf{x}}_2, \mathbf{y}_1, \bar{\mathbf{y}}_2] \end{array} \right] \quad (55)$$

etc.

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

1. Generate TSEM inequalities on the inflation DAG.
2. Factor the probabilities (same as Step 2 in ??).
3. Map as many probabilities to the original DAG as possible (same as Step 3 in ??).
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 [60] and Ghirardi and Marinatto [61]: Eq. (44) here corresponds to Eq. (2-4) in Ref. [60], and ?? here corresponds to Eq. (30) in Ref. [61].

IX. CONCLUSIONS

Our main contribution is a new way of deriving causal infeasibility criteria, namely the inflation DAG approach. An inflation DAG naturally implies nonlinear hybrid inequalities, i.e. containing gedankenprobabilities, which implicitly constrain the set of distribution consistent with the original causal structure. If desirable, one can further eliminate the gedankenprobabilities via quantifier elimination. Polynomial inequalities can be obtained through linear elimination techniques.

These inequalities are necessary but not sufficient conditions on a joint distribution for it to be explained by the causal structure. We've seen that they can be stronger than entropic inequalities sometimes, yet just how strong they are is still unclear. A distribution might satisfy all our polynomial inequalities and yet not be realizable from the causal structure. Our methods yields tight causal infeasibility criteria for Bell scenarios, but those scenarios are exceptional in that the sets of realizable distributions form a convex polytope.

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-broadcasting theorem [23, 24] on its head by emphasizing that classical hidden 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, 20, 25]. 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 amount by which a linear inequality [62, 63] is violated in quantum theory, but finding the quantum violation of a *polynomial* inequality is a more challenging task [64].

The difference between classical ontic-state duplication and quantum no-broadcasting makes the inequalities that result from our consideration to be especially suited for distinguishing the set of quantum-realizable distributions from its subset of classically-realizable 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, 20], 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.

Appendix A: Nonlinear Quantifier Elimination Approach

In the main text we considered deriving constraints on the pre-injectable sets via the marginal problem. There are further implications in the inflation DAG, however, which can be leveraged to further constrain the marginal distributions consistent with the inflation DAG.

Conditional Independence Relations

The most familiar such further constraints are the conditional independence relations relating observable variables in the inflation DAG. Conditional independence relations are inferred by d -separation; if \mathbf{X} and \mathbf{Y} are d -separated in the (inflation) DAG by \mathbf{Z} , then we infer $\mathbf{X} \perp \mathbf{Y} | \mathbf{Z}$. The d -separation criterion is explained at length in Refs. [1, 3, 5, 7], we present it only informally here.

We first define the notion of **exclusive ancestors**, because the usual notion of ancestors is inclusive, i.e. $\text{An}(\mathbf{Z})$ contains \mathbf{Z} . Let the exclusive ancestors of \mathbf{Z} so that $\text{An}(\setminus \mathbf{Z}) := \text{An}(\mathbf{Z}) \setminus \mathbf{Z}$. Next we define **conditioning a DAG** over \mathbf{Z} , i.e. $G_{\setminus \mathbf{Z}}$, which is formed by deleting any outgoing edges from \mathbf{Z} in G . Conditional DAGs enable the definition of **conditional ancestors**, i.e.

$$\text{An}_G(\mathbf{X} | \mathbf{Z}) := \begin{cases} \text{An}_{G_{\setminus \mathbf{Z}}}(\mathbf{X}) & \text{if } \text{An}_{G_{\setminus \mathbf{Z}}}(\mathbf{X}) \text{ and } \text{An}_G(\setminus \mathbf{Z}) \text{ are disjoint} \\ \text{An}_{G_{\setminus \mathbf{Z}}}(\mathbf{X}) \cup \text{An}_G(\setminus \mathbf{Z}) & \text{if } \text{An}_{G_{\setminus \mathbf{Z}}}(\mathbf{X}) \text{ and } \text{An}_G(\setminus \mathbf{Z}) \text{ overlap.} \end{cases} \quad (\text{A-1})$$

Finally, \mathbf{X} and \mathbf{Y} are d -separated by \mathbf{Z} iff $\text{An}_G(\mathbf{X} | \mathbf{Z})$ and $\text{An}_G(\mathbf{Y} | \mathbf{Z})$ are disjoint.

Every conditional independence relation can be translated into a nonlinear constraint on probabilities, as $\mathbf{X} \perp \mathbf{Y} | \mathbf{Z}$ implies $p(\mathbf{x}\mathbf{y} | \mathbf{z}) = p(\mathbf{x} | \mathbf{z})p(\mathbf{y} | \mathbf{z})$ for all \mathbf{x} , \mathbf{y} , and \mathbf{z} . As we generally prefer to work with unconditional probabilities, we summarize as follows: If \mathbf{X} and \mathbf{Y} are d -separated by \mathbf{Z} then $p(\mathbf{x}\mathbf{y}\mathbf{z})p(\mathbf{z}) = p(\mathbf{x}\mathbf{z})p(\mathbf{y}\mathbf{z})$ for all \mathbf{x} , \mathbf{y} , and \mathbf{z} . Such nonlinear constraints can be incorporated as restrictions on distributions consistent with the inflation DAG, supplementing the more trivial marginal-problem type constraints, and hence implicitly further constraining the injectable sets. Nonlinear quantifier elimination can then be used to make the implied constraints explicit.

Many modern computer algebra systems have functions capable of tackling nonlinear quantifier elimination symbolically⁷. Currently, however, it is not practical to perform nonlinear quantifier elimination on large polynomial systems with many quantifiers. Nevertheless, the nonlinear constraints can be easily accounted for numerically. Upon substituting numeric values for all the injectable probabilities the former quantifier elimination problem is converted to a universal quantifier existence problem: Do there exist gedankenprobabilities that satisfy the full set of linear and nonlinear constraints? Most computer algebra systems can resolve such *satisfiability* questions quite rapidly⁸.

TABLE I. A comparison of different approaches constraining the pre-injectable sets. The primary divide is quantifier elimination versus quantifier existence, with approaches being further subdivided into linear and nonlinear variants.

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 Partial Marginal Problem	Easy	Necessary-but-not-Sufficient Polynomial Inequalities

It is also possible to use a mixed strategy of linear and nonlinear quantifier elimination, such as ChavesPolynomial Chaves [66] advocates. The explicit results of Ref. [66] are therefore consequences of any inflation DAG, achieved by applying a mixed quantifier elimination strategy.

Coinciding Marginal Distributions

In principle one can exploit any d -separation criteria from the inflation DAG in order to transform or supplement the inequalities. If \mathbf{X} and \mathbf{Y} are d -separated by Z , then $p(\mathbf{xyz}) = \frac{p(\mathbf{xz})p(\mathbf{yz})}{p(\mathbf{z})}$ [1-4].

At the moment this is just a text dump from the earlier manuscript. Needs extensive cleanup!!!

These are the kinds of equalities and inequalities we can make use of. Move to last? These are sort of optional - we don't actually use them... Just as ancestral independence of nodes implies marginal independence of variables, so too d -separation of nodes implies conditional independence. If \mathbf{X} and \mathbf{Y} are don't share a common ancestor, then $\forall_{xy} p(\mathbf{xy}) = p(\mathbf{x})p(\mathbf{y})$. If \mathbf{X} and \mathbf{Y} are d -separated by Z , then $\forall_{xyz} p(\mathbf{xyz})p(\mathbf{z}) = p(\mathbf{xz})p(\mathbf{yz})$ [1-4]. For example, in Fig. 3

$$\begin{array}{l}
 A_2 \not\sim \{B_1, C_2\} \\
 B_2 \not\sim \{A_2, C_1\} \\
 C_2 \not\sim \{A_1, B_2\}
 \end{array}
 \quad \text{implies that} \quad
 \begin{array}{l}
 P(A_1 A_2 B_2 C_2) = P(C_2) \times P(A_1 A_2 B_2) \\
 P(A_2 B_1 B_2 C_2) = P(A_2) \times P(B_1 B_2 C_2) \\
 P(A_2 B_2 C_1 C_2) = P(B_2) \times P(A_2 C_1 C_2) \\
 P(A_2 B_2 C_2) = P(A_2) \times P(B_2) \times P(C_2)
 \end{array}
 \quad (\text{A-2}) \quad [\text{eqs}] \{\text{eqs}\}$$

To be clear, each equality in Eqs. (A-2) implies many more equalities at the level of individual probabilities, such as $p(a_2 b_1 b_2 c_2) = p(a_2)p(b_1 b_2 c_2)$, $p(a_2 b_1 b_2) = p(a_2)p(b_1 b_2)$, $p(a_2 b_1) = p(a_2)p(b_1)$, etc.

Deriving polynomial inequalities can be done in four steps.

1. Construct linear non-negativity inequalities.
2. Convert the inequalities to nonlinear ones via conditional independence relations.
3. Use injectable sets to connect the inequalities to the original DAG.
4. Perform quantifier elimination.

Step 1: Generate an initial set of linear inequalities

TODO: Change to EQUATIONS and inequalities, so that the inequality-only picture follows from the original one... UNDERWAY. EDITS IN PROGRESS HERE.

⁷ For example *Mathematica*TM's `Resolve` command, *Redlog*'s `rlposqe`, or *Maple*TM's `RepresentingQuantifierFreeFormula`, etc. BarFT-SMTLIB

⁸ For example *Mathematica*TM `ReduceExistsRealQ` function. Specialized satisfiability software such as SMT-LIB's `check-sat` [65] are particularly apt for this purpose. One can also exploit the fact that any nonlinear optimizer will return an error 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*

In order to derive inequalities on the original DAG, we begin by composing a set of linear conditions pertaining to the observable variables in the inflated DAG. Let the initial set be the nonnegativity of probability for all possible assignments to all observable variables. 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 $0 \leq P(A_1 A_2 B_1 B_2 C_1 C_2)$. The starting nonnegativity inequalities therefore look like

$$\begin{aligned}
0 &\leq p(a_1 a_2 b_1 b_2 c_1 c_2) \\
0 &\leq [p(\bar{a}_1 a_2 b_1 b_2 c_1 c_2) =] p(a_2 b_1 b_2 c_1 c_2) - p(a_1 a_2 b_1 b_2 c_1 c_2) \\
0 &\leq [p(a_1 \bar{a}_2 b_1 b_2 c_1 c_2) =] p(a_1 b_1 b_2 c_1 c_2) - p(a_1 a_2 b_1 b_2 c_1 c_2) \\
&\vdots \\
0 &\leq [p(\bar{a}_1 \bar{a}_2 b_1 b_2 c_1 c_2) =] p(b_1 b_2 c_1 c_2) - p(a_1 b_1 b_2 c_1 c_2) - p(a_2 b_1 b_2 c_1 c_2) + p(a_1 a_2 b_1 b_2 c_1 c_2) \\
0 &\leq [p(\bar{a}_1 a_2 \bar{b}_1 b_2 c_1 c_2) =] p(a_2 b_2 c_1 c_2) - p(a_1 a_2 b_2 c_1 c_2) - p(a_2 b_1 b_2 c_1 c_2) + p(a_1 a_2 b_1 b_2 c_1 c_2) \\
&\vdots
\end{aligned} \tag{A-3}$$

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

Step 2: Infer factorization of probabilities from structural independence

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 ancestral independences. In Fig. 3, for example, there are six pairs of ancestrally independent individual nodes, which consequently imply many marginal independences.

$$\begin{array}{ll}
\begin{array}{l} A_2 \not\perp\!\!\!\perp \{B_1, C_2\} \\ B_2 \not\perp\!\!\!\perp \{A_2, C_1\} \\ C_2 \not\perp\!\!\!\perp \{A_1, B_2\} \end{array} & \text{implies that} \quad \begin{array}{l} P(A_1 A_2 B_2 C_2) = P(C_2) \times P(A_1 A_2 B_2) \\ P(A_2 B_1 B_2 C_2) = P(A_2) \times P(B_1 B_2 C_2) \\ P(A_2 B_2 C_1 C_2) = P(B_2) \times P(A_2 C_1 C_2) \\ P(A_2 B_2 C_2) = P(A_2) \times P(B_2) \times P(C_2) \end{array}
\end{array} \tag{A-4}$$

etc. Interestingly, the original DAG in Fig. 1 does not imply any observable marginal independence relations, nor any observable conditional independence relations at all, for that matter.

By accounting for marginal independence relations, the set of linear inequalities is transformed into a set of polynomial inequalities. Eqs. (A-4) imply factorization of many different probabilities, such as $p(a_2 b_1 b_2 c_2) = p(a_2) p(b_1 b_2 c_2)$, $p(a_2 b_1 b_2) = p(a_2) p(b_1 b_2)$, $p(a_2 b_1) = p(a_2) p(b_1)$, etc. In principle one can exploit any d -separation criteria from the inflation DAG in order to transform or supplement the inequalities. If \mathbf{X} and \mathbf{Y} are d -separated by Z , then $p(\mathbf{x}\mathbf{y}\mathbf{z}) = \frac{p(\mathbf{x}\mathbf{z})p(\mathbf{y}\mathbf{z})}{p(\mathbf{z})}$ [1–4]. For simplicity, however, we have restricted our attention here to exclusively ancestral independences.

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

Here we account for injectable sets. As any subset of an injectable set is also an injectable set, we enumerate here only *maximal* injectable sets. The maximal injectable sets in Fig. 3 are

$$\begin{array}{ll}
\begin{array}{l} \{A_1 B_2\}, \\ \{B_1 C_2\}, \\ \{A_2 C_1\}, \\ \text{and } \{A_1 B_1 C_1\}, \end{array} & \text{which in turn imply that} \quad \begin{array}{l} P(A_1 B_2) = P(AB), \\ P(B_1 C_2) = P(BC), \\ P(A_2 C_1) = P(AC), \\ \text{and } P(A_1 B_1 C_1) = P(ABC). \end{array}
\end{array} \tag{A-5}$$

The injectable sets allow us to transform our set of inequalities once more, this time into inequalities which have bearing on the original causal structure. We call this transformation of probabilities the **injection map**; the injection map pursuant to Eqs. (A-5) takes

$$\begin{array}{ll}
\begin{array}{l} \{p(a_1) = p(a_2)\} \rightarrow p(a) \\ \{p(b_1) = p(b_2)\} \rightarrow p(b) \\ \{p(c_1) = p(c_2)\} \rightarrow p(c) \end{array} & \begin{array}{l} \{p(a_1 b_1) = p(a_1 b_2)\} \rightarrow p(ab) \\ \{p(a_1 c_1) = p(a_2 c_1)\} \rightarrow p(ac) \\ \{p(b_1 c_1) = p(b_1 c_2)\} \rightarrow p(bc) \\ p(a_1 b_1 c_1) \rightarrow p(abc) \end{array}
\end{array} \tag{A-6}$$

By the *definition* of inflation. Via the injectable sets.

After applying the injection map we are left with 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 appear alongside many probabilities which *do not inject* 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 $\{p(a_1a_2), p(a_1a_2b_1), \dots\}$. The probabilities pertaining to the inflated DAG which have no parallel in the original DAG are precisely probabilities regarding non-injectable sets of variables.

We name these non-injectable probabilities **gedankenprobabilities**, as they could be measured in-principle if one were to physically construct the inflated causal structure⁹. As we are really only concerned with the original DAG, however, these “unmeasured” 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.

As any superset of a non-injectable set is also not injectable, we enumerate here only *minimal* non-injectable sets. The *minimal* non-injectable sets in Fig. 3 are

$$\{A_1A_2\}, \{B_1B_2\}, \{C_1C_2\}, \{A_1B_1C_2\}, \{A_1B_2C_1\}, \{A_2B_1C_1\}. \quad (\text{A-7}) \quad \boxed{\text{eq:corei}}$$

Eq. (A-7) implies that our set of hybrid inequalities will be riddled with 34 different gedankenprobabilities.

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 essentially a form of quantifier elimination, where the quantifiers to be eliminated are the gedankenprobabilities. 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 related nonlinearly.

Many modern computer algebra systems have functions capable of tackling this sort of problem fully symbolically¹⁰. 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 injectable probabilities appearing 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 resulting system of polynomial inequalities? Most computer algebra systems can resolve such *satisfiability* questions quite rapidly¹¹.

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 given *value*, 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.

Appendix B: Further Polytope Projection Algorithms

The Equality Set Projection (ESP) algorithm [42, 43] 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.

⁹ The inflated causal structure is always hypothetically constructable, classically, hence the thought-experiment terminology. A gedankenprobability is also a well-defined hypothetical in quantum theory whenever the variables comprise a non-broadcasting set.

¹⁰ For example *Mathematica*TM’s **Resolve** command, *Redlog*’s **rlposqe**, or *Maple*TM’s **RepresentingQuantifierFreeFormula**, etc. [BarFT-SMTLIB](#)

¹¹ For example *Mathematica*TM’s **Reduce`ExistsRealQ** function. Specialized satisfiability software such as SMT-LIB’s **check-sat** [65] are particularly apt for this purpose. One can also exploit the fact that any nonlinear optimizer will return an error 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**.

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 [49]. 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 $\{\vec{x}, \vec{y} | \hat{A} \cdot \vec{x} + \hat{B} \cdot \vec{y} \geq \vec{c}\}$, where y are the quantifiers. If we can find any completely nonnegative vector \mathbf{w} such that $\mathbf{w} \cdot \hat{B} = \vec{0}$ then we automatically establish the quantifier-free inequality $\mathbf{w} \cdot \hat{A} \cdot \vec{x} \geq \mathbf{w} \cdot \vec{c}$. Solving for “random” nonnegative vectors \mathbf{w} is easy; solving for all possible solutions is rather more difficult. Balas [67] refined this method so that each extremal construction of \mathbf{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 C: 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 45]¹². This is related to the V-rep characterization of polyhedral cones: If a cone is defined such that $W_{\hat{M}} := \{\vec{x} | \exists \mathbf{v} \geq \mathbf{0} : \hat{M} \cdot \mathbf{v} = \vec{x}\}$ then $\vec{b} \in W_{\hat{M}}$ if and only if the linear system of equations $\hat{M} \cdot \mathbf{v} = \vec{b}$ has a solution such that all the elements of \mathbf{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*TM 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

¹² The “if” is obvious. The “only if” is a consequence of Farka’s lemma [45].

efficient low-level code and *Mathematica*TM'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}];
InternalHandlerBlock[{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:

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

If \hat{A} is the matrix whose 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.

Appendix D: Recognizing observationally equivalent DAGs

One expects that an edge $A \rightarrow B$ 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 \mathbf{Z} , hereafter $\mathbf{Z} \models X$, if X can be completely inferred from knowledge of \mathbf{Z} alone. In a deterministic DAG, for example, every non-root node is perfectly predictable given its parents, $\text{pa}[X] \models X$. Indeed, in a deterministic DAG the node X is perfectly predictable from \mathbf{Z} if X is a deterministic descendant of \mathbf{Z} . Operationally, X is a deterministic descendant of \mathbf{Z} if the intersection of [the ancestors of X] with [the non-ancestors of \mathbf{Z}] is a subset of [the descendants of \mathbf{Z}]. Happily though, perfect predictability can be extrapolated from a causal structure with minimal effort: $\mathbf{Z} \models X$ if every directed path to X from any root node is blocked by \mathbf{Z} .

Markov Blanket: The Markov Blanket for a set of nodes \mathbf{V} , hereafter $\text{MB}[\mathbf{V}]$, is the set of all of \mathbf{V} 's children, parents, and co-parents. The Markov Blanket is so defined because the nodes in \mathbf{V} are conditionally independent of *everything* given $\text{MB}[\mathbf{V}]$. If the random variables in the Markov Blanket $\text{MB}[\mathbf{V}]$ are known, then information about nodes inside \mathbf{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 \mathbf{Z} is a Markov Partition for a pair of random variables X and Y , hereafter $X \dashv \mathbf{Z} \vdash Y$, if the pair are conditionally independent of each other given *any superset* of \mathbf{Z} . Operationally, this means that X and Y are d -separated by every superset of \mathbf{Z} . Equivalently, $X \dashv \mathbf{Z} \vdash Y$ if $\text{MB}[\mathbf{V}] \subseteq \mathbf{Z}$ and $X \in \mathbf{V}$ while $Y \notin \mathbf{V}$, or if $\text{MB}[\mathbf{V}] \subseteq \mathbf{Z}$ and $Y \in \mathbf{V}$ while $X \notin \mathbf{V}$. Happily though, Markov Partitions can be extrapolated from a causal structure with minimal effort: $X \dashv \mathbf{Z} \vdash Y$ if and only if X and Y would be in *disconnected components* under the deletion of all edges initiation from \mathbf{Z} .

Sufficient Statistic: A set of nodes \mathbf{Z} is a sufficient statistic for A relative to X , hereafter $\mathbf{Z} \vdash A|X$, 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 \mathbf{Z} instead. In other words, learning A can never teach anything new about X if \mathbf{Z} is already known. If $X = A$, then the *only way* \mathbf{Z} can stand in for A when making inferences about A is if A is perfectly predicable given \mathbf{Z} , i.e. $\mathbf{Z} \vdash A|A \iff \mathbf{Z} \models A$. If $A \neq B$ then there are four *and only four?* ways that $\mathbf{Z} \vdash A|X$ can be implied by a DAG: If $\mathbf{Z} \models A$, if $\mathbf{Z} \models X$, if $\text{MB}[\mathbf{V}] \subseteq \mathbf{Z}$ and $A \in \mathbf{V}$ while $X \notin \mathbf{V}$, or if $\text{MB}[\mathbf{V}] \subseteq \mathbf{Z}$ and $X \in \mathbf{V}$ while $A \notin \mathbf{V}$. *Alternatively:* If $A \neq B$ then there are THREE ways that $\mathbf{Z} \vdash A|X$ can be implied by a DAG: If $\mathbf{Z} \models A$, if $\mathbf{Z} \models X$, and if $A \dashv \mathbf{Z} \vdash X$.

edgeadding

Theorem 2. An edge $A \rightarrow B$ can be added to G without observational impact if $\text{pa}[B]$ are a sufficient statistic for A relative to all observable nodes, i.e. $\forall_{\text{observable } X} : \text{pa}[B] \vdash A|X$.

In particular, the edge $A \rightarrow B$ can always be added whenever $\text{pa}[B] \models A$, including, but not limited to, the instance $\text{pa}[A] \subseteq \text{pa}[B]$.

Furthermore, the edge $\Lambda \rightarrow B$ can be also always be added whenever Λ is latent and $\text{MB}[\Lambda] \subseteq \text{pa}[B]$.

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

edropping

Corollary 2.1. An edge $A \rightarrow B$ can be dropped from G to form G' such that G and G' are observationally equivalent if *and only if* the edge $A \rightarrow B$ can be added (back) to G' while leaving G' observationally invariant per Theorem 2.

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

entadding

Theorem 3. A (root) latent node Λ can be removed from G without observational impact if Λ has only one child node and no co-parents (Λ is “equivalent to local randomness”), or if Λ 's children are also all children of another single latent node (Λ is “covered-for by another latent node”). Conversely, a new root latent node Λ can be introduced along with various outgoing edges, without observational impact, if Λ would be equivalent to local randomness or covered-for by another latent node.

Naturally, two causal structures are observationally equivalent if one can be transformed into the other without observational impact, via Theorems 2 and 3. Some examples of observationally equivalent scenarios, and the steps which interconvert them, are given in Fig. 12.

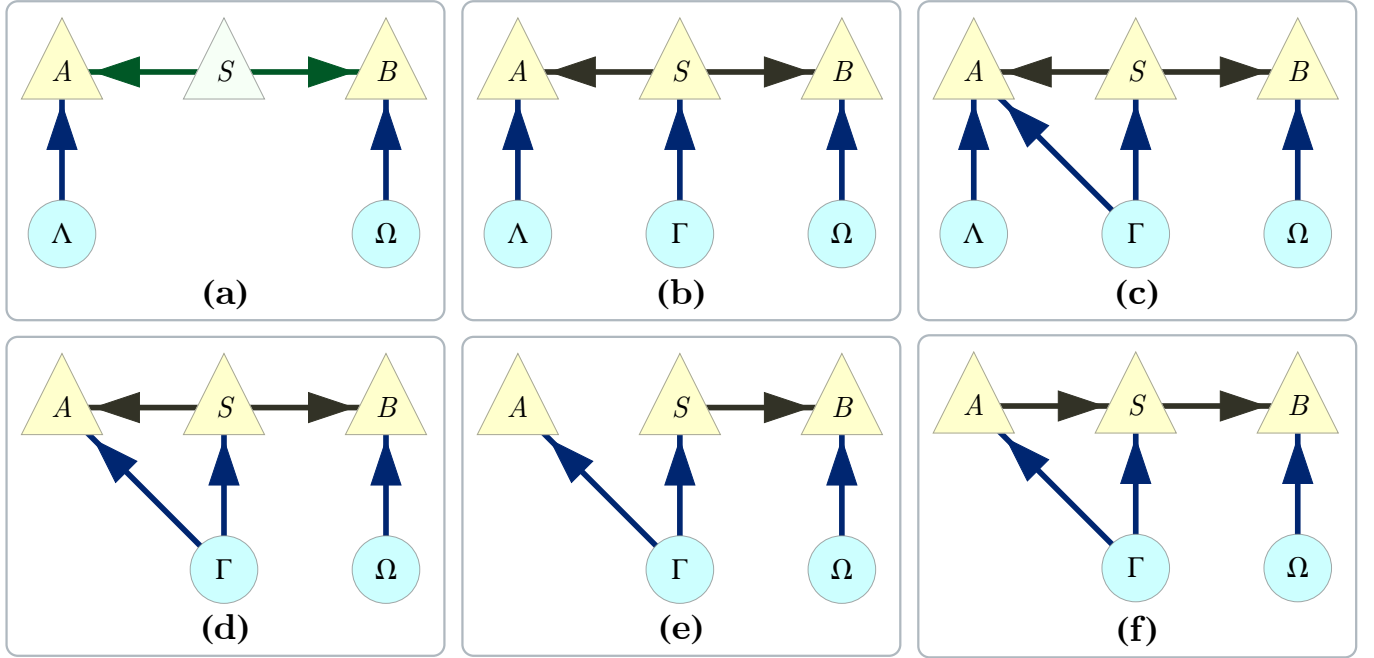


FIG. 12. A set of observational equivalent causal structures. The reasons the changes are observational invariant are as follows:
(a)~(b) because Γ is useless in (b), and as such Γ can be dropped from (b) per Theorem 3.
(b)~(c) because $\text{MB}[\Gamma] \subseteq \text{pa}[A]$ in (b), and as such $\Gamma \rightarrow A$ can be added to (b) per Theorem 2.
(c)~(d) because Λ is redundant to Γ in (c), and as such Λ can be dropped from (c) per Theorem 3.
(d)~(e) because $\text{pa}[S] \subseteq \text{pa}[A]$ in (e), and as such $S \rightarrow A$ can be added to (e) per Theorem 2.
(e)~(f) because $\text{pa}[A] \subseteq \text{pa}[S]$ in (e), and as such $A \rightarrow S$ can be added to (e) per Theorem 2.

fig:equiv

Appendix E: 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 4. *The following inequalities hold for all classical correlations in the triangle scenario:*

- (a) $p(a)p(c) \leq p(ab) + p(\bar{b}c)$
- (b) $p(ab\bar{c})p(\bar{a}bc)p(\bar{a}bc) \leq p(abc) + p(\bar{a}\bar{b})p(ab\bar{c})p(a) + p(\bar{a}\bar{c})p(\bar{a}bc)p(c) + p(\bar{b}\bar{c})p(\bar{a}bc)p(b)$
- (c) $p(ab\bar{c})p(\bar{a}bc)p(\bar{a}bc) \leq p(abc)^2 + 2(p(\bar{a}\bar{b})p(ab\bar{c}) + p(\bar{a}\bar{c})p(\bar{a}bc) + p(\bar{b}\bar{c})p(\bar{a}bc))$
- (d) $p(abc)^2p(\bar{a}\bar{b}\bar{c}) \leq p(ab\bar{c})p(\bar{a}bc)p(\bar{a}bc) + (2p(abc) + p(\bar{a}\bar{b}\bar{c}))(1 - p(abc) - p(\bar{a}\bar{b}\bar{c}))$
- (e) $p(abc)^2p(\bar{a}\bar{b}\bar{c}) \leq p(abc)^3 + (2p(abc) + p(\bar{a}\bar{b}\bar{c}))(1 - p(abc) - p(\bar{a}\bar{b}\bar{c}))$
- (f) $p(a)p(b)p(c) \leq p(\bar{a}\bar{b}\bar{c}) + p(ab)p(c) + p(ac)p(b) + p(bc)p(a)$
- (g) $p(a)p(b)p(c) \leq p(\bar{a}\bar{b}\bar{c})^2 + 2(p(ab)p(c) + p(ac)p(b) + p(bc)p(a))$

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

[illegible]

-
- | |
|------------|
| causality |
| causation |
| abilistic |
| abilistic |
| dSpekkens |
| z2012bell |
| y2014gdag |
| ondBellIII |
| retronym |
| 3marginal |
| 2014novel |
| inference |
| 11966lhvm |
| etwork2012 |
| 012device |
| r2013Bell |
| IApproach |
| tzDuality |
| nalstates |
| foquantum |
| relations |
| ancestors |
| antum1996 |
| neral2006 |
| ignalling |
| 2entropic |
| 8entropic |
| nequality |
| 1989going |
| bits2Ways |
| 6entropic |
| ovariance |
| 4einstei |
| HOriginal |
- [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] Dictionary.com, “Retronym,” (2016).
 - [10] T. Fritz and R. Chaves, “Entropic inequalities and marginal problems,” *IEEE Trans. Info. Theo.* **59**, 803 (2013).
 - [11] R. Chaves, L. Luft, and D. Gross, “Causal structures from entropic information: geometry and novel scenarios,” *New J. Phys.* **16**, 043001 (2014).
 - [12] 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.
 - [13] J. S. Bell, “On the problem of hidden variables in quantum mechanics,” *Rev. Mod. Phys.* **38**, 447 (1966).
 - [14] 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).
 - [15] V. Scarani, “The device-independent outlook on quantum physics,” *Acta Physica Slovaca* **62**, 347 (2012).
 - [16] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, “Bell nonlocality,” *Rev. Mod. Phys.* **86**, 419 (2014).
 - [17] J.-D. Bancal, *On the Device-Independent Approach to Quantum Physics* (Springer International Publishing, 2014).
 - [18] T. Fritz, “Polyhedral duality in Bell scenarios with two binary observables,” *J. Math. Phys.* **53**, 072202 (2012), 10.1063/1.4734586.
 - [19] M. S. Leifer and R. W. Spekkens, “Towards a formulation of quantum theory as a causally neutral theory of Bayesian inference,” *Phys. Rev. A* **88**, 052130 (2013).
 - [20] R. Chaves, C. Majenz, and D. Gross, “Information-theoretic implications of quantum causal structures,” *Nat. Comm.* **6**, 5766 (2015).
 - [21] C. Branciard, D. Rosset, N. Gisin, and S. Pironio, “Bilocal versus nonbilocal correlations in entanglement-swapping experiments,” *Phys. Rev. A* **85**, 032119 (2012).
 - [22] B. Steudel and N. Ay, “Information-theoretic inference of common ancestors,” *Entropy* **17**, 2304 (2015).
 - [23] 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).
 - [24] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, “Cloning and broadcasting in generic probabilistic theories,” *quant-ph/0611295* (2006).
 - [25] R. Chaves and C. Budroni, “Entropic nonsignalling correlations,” *arXiv:1601.07555* (2015).
 - [26] R. Chaves and T. Fritz, “Entropic approach to local realism and noncontextuality,” *Phys. Rev. A* **85**, 032113 (2012).
 - [27] S. L. Braunstein and C. M. Caves, “Information-theoretic Bell inequalities,” *Phys. Rev. Lett.* **61**, 662 (1988).
 - [28] B. W. Schumacher, “Information and quantum nonseparability,” *Phys. Rev. A* **44**, 7047 (1991).
 - [29] D. M. Greenberger, M. A. Horne, and A. Zeilinger, “Going beyond bell’s theorem,” in *Bell’s theorem, quantum theory and conceptions of the universe*, edited by M. Kafatos (Springer, 1989) pp. 69–72.
 - [30] W. Dür, G. Vidal, and J. I. Cirac, “Three qubits can be entangled in two inequivalent ways,” *Phys. Rev. A* **62**, 062314 (2000).
 - [31] M. Weilenmann and R. Colbeck, “Analysis of the entropic approach to distinguish classical and quantum causal structures,” personal communication (2016).
 - [32] J. Aberg, R. Chaves, D. Gross, A. Kela, and K. U. von Prillwitz, “Inferring causal structures with covariance information,” poster session, Quantum Networks conference (2016).
 - [33] J. S. Bell, “On the Einstein-Podolsky-Rosen paradox,” *Physics* **1**, 195 (1964).
 - [34] 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).

- convexity [35] J. M. Donohue and E. Wolfe, “Identifying nonconvexity in the sets of limited-dimension quantum correlations,” *Phys. Rev. A* **92**, 062120 (2015).
- relaxation [36] 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.
- ROriginal [37] S. Popescu and D. Rohrlich, “Quantum nonlocality as an axiom,” *Found. Phys.* **24**, 379 (1994).
- PRUnit [38] J. Barrett and S. Pironio, “Popescu-rohrlich correlations as a unit of nonlocality,” *Phys. Rev. Lett.* **95**, 140401 (2005).
- entingCRA [39] K. Korovin, N. Tsiskaridze, and A. Voronkov, “Implementing conflict resolution,” *Perspectives of Systems Informatics*, **362** (2012).
- implexSAT [40] F. Bobot, S. Conchon, E. Contejean, M. Iguernelala, A. Mahboubi, A. Mebsout, and G. Melquiond, “A simplex-based extension of fourier-motzkin for solving linear integer arithmetic,” *Automated Reasoning*, **67** (2012).
- neTheorem [41] A. Fine, “Hidden variables, joint probability, and the Bell inequalities,” *Phys. Rev. Lett.* **48**, 291 (1982).
- equality [42] 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).
- thesis2005 [43] C. Jones, *Polyhedral Tools for Control*, Ph.D. thesis, University of Cambridge (2005).
- Jones2008 [44] C. N. Jones, E. C. Kerrigan, and J. M. Maciejowski, “On polyhedral projection and parametric programming,” *J. Optimiz. Theo. Applic.* **138**, 207 (2008).
- projection [45] A. Fordan, *Projection in Constraint Logic Programming* (Ios Press, 1999).
- tzigEaves [46] G. B. Dantzig and B. C. Eaves, “Fourier-Motzkin elimination and its dual,” *J. Combin. Th. A* **14**, 288 (1973).
- shapot2012 [47] D. V. Shapot and A. M. Lukatskii, “Solution building for arbitrary system of linear inequalities in an explicit form,” *Am. J. Comp. Math.* **02**, 1 (2012).
- strakov2015 [48] S. I. Bastrakov and N. Y. Zolotykh, “Fast method for verifying Chernikov rules in Fourier-Motzkin elimination,” *Comp. Mat. & Math. Phys.* **55**, 160 (2015).
- is2000lrs [49] D. Avis, “A revised implementation of the reverse search vertex enumeration algorithm,” *Polytopes — Combinatorics and Computation*, **177** (2000).
- qskeleton [50] S. Bastrakov, “qskeleton,” sbastrakov.github.io/qskeleton (2016), polyhedral computation software for Fourier-Motzkin elimination with Chernikov rules.
- rNetworks [51] A. Tavakoli, P. Skrzypczyk, D. Cavalcanti, and A. Acín, “Nonlocal correlations in the star-network configuration,” *Phys. Rev. A* **90**, 062109 (2014).
- tNetworks [52] 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).
- cNetworks [53] A. Tavakoli, “Bell-type inequalities for arbitrary noncyclic networks,” *Phys. Rev. A* **93**, 030101 (2016).
- nstraints [54] 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.
- ToyTheory [55] R. W. Spekkens, “Evidence for the epistemic view of quantum states: A toy theory,” *Phys. Rev. A* **75**, 032110 (2007).
- um2012GPT [56] H. Barnum and A. Wilce, “Post-classical probability theory,” [arXiv:1205.3833](https://arxiv.org/abs/1205.3833) (2012).
- ta2014GPT [57] P. Janotta and H. Hinrichsen, “Generalized probability theories: what determines the structure of quantum theory?” *J. Phys. A* **47**, 323001 (2014).
- oecke2011 [58] B. Coecke and R. W. Spekkens, “Picturing classical and quantum Bayesian inference,” *Synthese* **186**, 651 (2011).
- elson1980 [59] B. S. Cirel’son, “Quantum generalizations of Bell’s inequality,” *Lett. Math. Phys.* **4**, 93 (1980).
- owsky1989 [60] I. Pitowsky, *Quantum Probability - Quantum Logic*, Lecture Notes in Physics, Berlin Springer Verlag, Vol. 321 (Springer Berlin Heidelberg, 1989).
- hirardi08 [61] G. Ghirardi and L. Marinatto, “Proofs of nonlocality without inequalities revisited,” *Phys. Lett. A* **372**, 1982 (2008).
- A2008Long [62] 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).
- I3322NPA1 [63] K. F. Pál and T. Vértesi, “Quantum bounds on bell inequalities,” *Phys. Rev. A* **79**, 022120 (2009).
- NPARreview [64] S. Pironio, M. Navascués, and A. Acín, “Convergent relaxations of polynomial optimization problems with noncommuting variables,” *SIAM J. Optim.* **20**, 2157 (2010).
- FT-SMTLIB [65] C. Barrett, P. Fontaine, and C. Tinelli, “The Satisfiability Modulo Theories Library (SMT-LIB),” www.SMT-LIB.org (2016).
- polynomial [66] R. Chaves, “Polynomial Bell inequalities,” *Phys. Rev. Lett.* **116**, 010402 (2016).
- staben [67] E. Balas, “Projection with a minimal system of inequalities,” *Comp. Optimiz. Applic.* **10**, 189 (1998).