Causal Compatibility Inequalities Admitting of Quantum Violations in the Triangle Structure

Thomas C. Fraser[1,2,*] and Elie Wolfe[1]

[1] *Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada, N2L 2Y5*
[2] *University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1*

(Dated: April 7, 2017)

It has long been recognized that certain quantum correlations are often incompatible with particular assumption about classical causal structure. Such correlations are sometimes termed non-local, although they are more accurately termed non-classical, as the presence of correlations certifies the genuinely non-classical nature of a causal setup in a device-independent fashion. Whenever a quantum correlation violates a causal compatibility inequality (e.g. a Bell inequality) then the correlation is witnessed as non-classical. Any constraint which is satisfied by all correlations which *are* classically compatible with a given causal structure defines a causal compatibility criterion. Such criteria were recently were recently derived for the Triangle structure [arXiv:1609.00672] in the form of polynomial inequalities, begging the question: do any of those inequalities admit violation by quantum correlations? Numerical investigation suggests not, and we further conjecture that the set of correlations admitted by the classical Triangle structure is equivalent to the set of correlations admitted by its quantum generalization whenever the three observable variables are binary. Our main contribution in this work, however, is the derivation of new causal compatibility inequalities for the Triangle structure which *do* admit quantum violation. We conclude by considering the possibility of quantum resources potentially qualitatively different from those known previously.

* tcfraser@tcfraser.com

CONTENTS

I. INTRODUCTION

In recent decades, the technical utility of quantum mechanics has become abundantly clear. In the realm of computation, quantum algorithms — such as Shor's algorithm [1] and numerous others [2] — scale exponentially better than their classical counterparts. In the realm of secure communication, quantum protocols — a popular example being Quantum Key Distribution [3] — are able to provide privacy even against hypothetical adversaries with unlimited computational power, a desideratum which classical protocols are unable to fulfill. Throughout history, every aspect of quantum phenomena which fails to be emulated by classical resources has led to a practical exploitation of the non-classical feature in order to solve a some computational of communicational problem [4]. Motivated by past successes, a primal objective of modern quantum information theory is the discovery of new situations wherein quantum mechanics offers an advantage, and certification that the quantum advantage is genuine.

From a foundational prospective, the most robust demonstrations of quantum phenomena with no classical emulation have involved Bell inequalities [5]. Originally, Bell inequalities were derived as a way to show that no hidden variable theory could ever account for quantum mechanics; in this sense Bell inequalities are a response to the famous EPR paradox [6]. The enumeration of Bell inequalities has since become a widespread systematic way of demonstration the *non-classicality* of a given observation. More recently it has been appreciated that Bell inequalities can understood as consequences of **causal inference** [7]. Causal inference is concerned with classifying observations into those which can and cannot be explained by a hypothesized causal structure. The abstract nature of causal inference is responsible for its presence in numerous scientific fields including machine learning and biology [8, 9]. Causal compatibility inequalities, such as Bell and Instrumental inequalities, characterize the space of observations that are compatible with a hypothesized causal structure, albeit the characterization offered by practically derivable inequalities is often only an approximation. To derive the traditional Bell inequalities from causal inference one starts with a (classical) causal structure known as the Bell structure, as depicted in Fig. 1. The fundamental Bell structure involves non-communicating parties making measurements on some hidden shared resource λ, where the measurement outcomes (A, B) are presumed to be stochastic functions of the the local choices of measurement settings (S_A, S_B) and the shared resource λ. Quantum non-classicality in the Bell structure has been thoroughly studied since Bell's original work [10]. More complex structures, however, such as the correlation scenarios proposed by Fritz [11, 12], are much less understood. Here we investigate one particular causal structure: the Triangle structure (Fig. 2).

The Triangle structure (Fig. 2) is a causal structure composed of 3 parties labeled A, B, C arranged in a triangular configuration while pair-wise sharing hidden/latent variables X, Y, Z. It has been extensively studied previously (see [13, Fig. 1], [14, Fig. 6], [15, Fig. 8], [16, Fig. 8, App. E], [11, Fig. 3], [17, Fig. 1], ...). An overview of some milestone results is provided in Section IV. Identifying causal compatibility inequalities for this configuration has been seen as particularly challenging [15]. Further identifying causal compatibility inequalities of such high resolution that such inequalities can be violated by quantum-accessible distributions has remained out-of-reach for the Triangle structure.

This work is the first, therefore, to find causal compatibility inequalities for the Triangle structure that are able to be violated by quantum-accessible distributions. This accomplishment was made possible through the combination of two previous developments. The first, an insight made by Fritz [11] regarding the ability to re-interpret the Bell structure as a portion of the Triangle structure. The second, a new framework for solving causal inference problems developed by Wolfe *et al.* [17] called **The Inflation Technique**. Ultimately, this work serves as a validation that the Inflation Technique is sensitive enough to offer new insights into quantum non-classicality. Moreover, these new inequalities offer an avenue for recognizing previously unknown forms of non-classicality. The authors attempts to find such novel resources were met with only partial success, suggesting clear refinements for future exploration.

The manuscript is organized into the following sections: Section II recalls important notions from causal inference theory and sets up the notation to be used. Section III offers a summary of the popular Bell structure and associated inequalities. Section IV discusses the Triangle structure and provides an overview of existing research; identifying its stark differences from the Bell structure and motivating why the Triangle structure was chosen as a case study. Section V defines and discusses a singularly-quantum correlation first conceived of by Fritz [11], which we term the **Fritz distribution**. The Fritz distribution was proven to be non-classical without the use of inequalities — this report fills in the desired missing inequalities. A sample of these inequalities are presented in Section VI (specifically Eq. (12) as well as in Eqs. (14,15)Todo (TC Fraser): Reorganize the location of the inequalities). Appendix A briefly summaries the Inflation Technique in the specific context of this work. Although the summary presented here is designed to be self-standing, a much more pedagogical introduction is offered by the original work [17]. Appendix A 3 demonstrates how the inflation technique can be used to derive causal compatibility inequalities for the Triangle structure that are violated by quantum-accessible distributions. Finally, Section VII introduces the method for parameterizing the space of quantum-accessible distributions which we used to numerically optimize against our derived inequalities in search of stronger types of non-classicality. Additionally, the Fritz distribution is subjected to noise in order to measure the robustness of the derived inequalities. Section VIII concludes with proposed avenues for further research.

II. CAUSAL COMPATIBILITY

The task of causal inference is to determine the set of potentially observable probability distributions compatible with some hypothesis about causal relationships [8]. If an observed distribution can be explained by the hypothesized causal mechanism, then the distribution is said to be *compatible* with said causal mechanism. In order to define compatibility rigorously, we first need to formally define the notion of a causal hypothesis.

A hypothesis of causal mechanism is formally referred to as a **causal structure** and can be represented as a **directed acyclic graph**. A directed graph $\mathcal{G}$ is an ordered tuple $(\mathcal{N}, \mathcal{E})$ of respectively *nodes* and *edges* where each edge $e \in \mathcal{E}$ connects a *pair* of nodes $n, m \in \mathcal{N}$ with a directed arrow $e = \{n \to m\}$. A directed graph is *acyclic* if there are no paths following the directions of the edges starting from and returning to the same node. The nodes $\mathcal{N}$ of a causal structure represent random variables while the edges $\mathcal{E}$ represent a casual influence from one variable to another pursuant to the prescribed direction.

Henceforth, we will utilize a number of familiar notions from graph theory and denote them accordingly. The **parents of a node** $n \in \mathcal{N}$ are all nodes which point directly into n, i.e. $\mathsf{Pa}_{\mathcal{G}}(n) \equiv \{m \mid m \to n\}$. Similarly defined are the **children of a node** $\mathsf{Ch}_{\mathcal{G}}(n) \equiv \{m \mid n \to m\}$. Recursively defined are the **ancestors of a node** $\mathsf{An}_{\mathcal{G}}(n) \equiv \bigcup_{i\in\mathbb{N}} \mathsf{Pa}^i_{\mathcal{G}}(n)$ where $\mathsf{Pa}^i_{\mathcal{G}}(n) \equiv \mathsf{Pa}_{\mathcal{G}}\big(\mathsf{Pa}^{i-1}_{\mathcal{G}}(n)\big)$ and $\mathsf{Pa}^0_{\mathcal{G}}(n) = n$ and the **descendants of a node** $\mathsf{De}_{\mathcal{G}}(n) \equiv \bigcup_{i\in\mathbb{N}} \mathsf{Ch}^i_{\mathcal{G}}(n)$ where $\mathsf{Ch}^i_{\mathcal{G}}(n) \equiv \mathsf{Ch}_{\mathcal{G}}\big(\mathsf{Ch}^{i-1}_{\mathcal{G}}(n)\big)$ and $\mathsf{Ch}^0_{\mathcal{G}}(n) = n$. Finally, we extend this notation to a subset of nodes $N \subseteq \mathcal{N}$ by performing a union over elements. As an example, the parents of the nodes $N \subseteq \mathcal{N}$ are denoted $\mathsf{Pa}_{\mathcal{G}}(N) = \bigcup_{n\in N} \mathsf{Pa}_{\mathcal{G}}(n)$.

Now that the notion of a causal structure $\mathcal{G}$ is defined along with some standard graph-theoretic notation, it is now possible to precisely define the notion of compatibility between $\mathcal{G}$ and a probability distribution $P_{\mathcal{N}}$ defined over the nodes of $\mathcal{G}$. Intuitively, the absence of certain causal connections within $\mathcal{G}$ should limit the set of possible observations that can be made on $\mathcal{G}$. Specifically, the causal structure $\mathcal{G}$ hypothesizes that each variable $n \in \mathcal{N}$ is causally influenced by its parents $\mathsf{Pa}_{\mathcal{G}}(n)$, then n should be independent of all other variables, except its own descendants when conditioned over $\mathsf{Pa}_{\mathcal{G}}(n)$. This is known as the **local Markov property**:[1]

$$n \perp\!\!\!\perp \big(\mathcal{N} \setminus \mathsf{De}_{\mathcal{G}}(n)\big) \mid \mathsf{Pa}_{\mathcal{G}}(n) \quad \forall n \in \mathcal{N} \tag{1}$$

If a given distribution $P_{\mathcal{N}}$ defined over *all* of the nodes $\mathcal{N}$ of $\mathcal{G}$ satisfies Eq. (1), then $P_{\mathcal{N}}$ is said to be **compatible with $\mathcal{G}$**. Alternatively yet equivalently, the notion of compatibility can be interpreted as follows: a distribution $P_{\mathcal{N}}$ is compatible with $\mathcal{G}$ if $P_{\mathcal{N}}$ can be consistently factorized into conditional probability distributions for each variable $n \in \mathcal{N}$ conditioned on its parents $\mathsf{Pa}_{\mathcal{G}}(n)$:

$$P_{\mathcal{N}} = \prod_{n \in \mathcal{N}} P_{n|\mathsf{Pa}_{\mathcal{G}}(n)} = P_{n_1|\mathsf{Pa}_{\mathcal{G}}(n_1)} \times \cdots \times P_{n_k|\mathsf{Pa}_{\mathcal{G}}(n_k)} \tag{2}$$

The conditional distributions in Eq. (2) (i.e. $\{P_{n|\mathsf{Pa}_{\mathcal{G}}(n)} \mid n \in \mathcal{N}\}$) are referred to as a set of **causal parameters for $\mathcal{G}$**. If a distribution $P_{\mathcal{N}}$ violates Eq. (1) or equivalently cannot be factorized according to Eq. (2), $P_{\mathcal{N}}$ is said to be **incompatible** with $\mathcal{G}$. When one is specified with a joint distribution $P_{\mathcal{N}}$ defined over all nodes of a causal structure $\mathcal{G}$, it is trivially easy to verify whether or not $P_{\mathcal{N}}$ is compatibility with $\mathcal{G}$; Appendix A 1 describes in detail how this is accomplished.

A challenge is present when one is supplied with a partial observation, i.e. a joint distribution $P_{\mathcal{N}_O}$ where $\mathcal{N}_O \subset \mathcal{N}$ is some subset of variables referred to as the **observable nodes** $\mathcal{N}_O$. In such cases, Eq. (2) can not be verified by direction calculation because it is certainly possible that the parents of some observable node are themselves unobservable $\mathsf{Pa}_{\mathcal{G}}(\mathcal{N}_O) \not\subseteq \mathcal{N}_O$. In such cases, compatibility between $\mathcal{G}$ and $P_{\mathcal{N}_O}$ depends on the existence or non-existence of a set of causal parameters for $\mathcal{G}$ such that $P_{\mathcal{N}_O} = \prod_{n \in \mathcal{N}_O} P_{n|\mathsf{Pa}_{\mathcal{G}}(n)}$. The unobservable nodes are termed **latent nodes** $\mathcal{N}_L = \mathcal{N} \setminus \mathcal{N}_O$ representing hidden random variables that are hidden either by some fundamental process or cannot be measured due to other limitations. Todo (TC Fraser): Explain why this is difficult more. Todo (TC Fraser): Figure out if I should include d-sep here Todo (TC Fraser): Explain what this work and inflation accomplishes regarding this

In practice however, checking causal parameters systemically is intractable, especially when the cardinality (number of outcomes) of the latent variables is unspecified. Instead, **causal compatibility inequalities**[2] are probabilistic inequalities defined over $P_{\mathcal{N}_O}$ and its induced marginals which constraint the set of compatible probability distributions. If a probability distribution $P_{\mathcal{N}_O}$ happens to *violate* any causal compatibility inequality, then that distribution deemed *incompatible*. In Appendix A 3, a brief discussion of existing techniques for deriving these inequalities will take place.

Unfortunately, a singular inequality cannot prove that a given distribution is *compatible*, only *incompatible*. Instead, a *complete characterization* of compatibility would consist of a complete set of all valid causal compatibility inequalities; satisfaction of a complete set of inequalities does prove compatibility. Currently however, it is unknown how to obtain a complete characterization for any given causal structure, including the Triangle structure.

From the perspective of identifying quantum non-classicality, a causal structure $\mathcal{G}$ adopts the role of a classical hypothesis. Therefore, non-classicality becomes synonymous with incompatibility: if a distribution $P_{\mathcal{N}_O}$ is incompatible, then it is non-classical. Henceforth, we will use these two terms interchangeably. From a resource standpoint, if the non-classicality of $P_{\mathcal{N}_O}$ can be witnessed by an inequality I, but $P_{\mathcal{N}_O}$ can be implemented using quantum states and measurements, then the causal compatibility inequality I represents a physical task where quantum resources out-perform classical resources.

III. BELL STRUCTURE

Todo (TC Fraser): Review of Bell structure, Tsirelson-box, CHSH inequality

$$\langle AB|S_A=0,S_B=0\rangle + \langle AB|S_A=0,S_B=1\rangle + \langle AB|S_A=1,S_B=0\rangle - \langle AB|S_A=1,S_B=1\rangle \leq 2 \tag{3}$$

---

[1] Here $A \perp\!\!\!\perp B \mid C$ denotes statistical independence between A and B conditional on C where $A, B, C \subseteq \mathcal{N}$, i.e. for a given probability distribution $P_{\mathcal{N}}$, $P_{AB|C} = P_{A|C}P_{B|C}$.

[2] We refer to these inequalities as "causal compatibility inequalities" instead of simply Bell inequalities for two reasons. Firstly, "Bell inequalities" usually are associated specifically with the Bell structure. Secondly, the inequalities derived in this work are fundamentally distinct from a typical Bell inequality in that these inequalities are *polynomial* over $P_{\mathcal{N}_O}$ instead of *linear*.

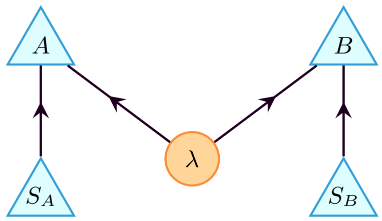


FIG. 1: The Bell structure consisting of two observers A, B together with measurement settings S_A and S_B respectively. The shared hidden variable is labeled λ.

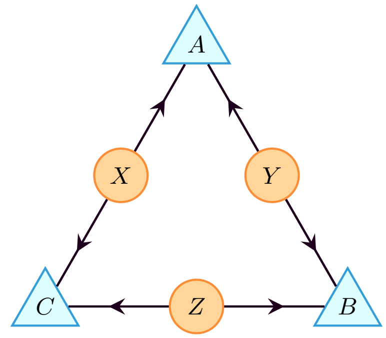


FIG. 2: The casual structure of the Triangle structure. Three variables A, B, C are observable, while X, Y, Z are latent variables.

IV. TRIANGLE STRUCTURE

Discovering quantum non-classicality in any causal structure is desirable, so why, in particular, is it important to do so in the Triangle structure? As was mentioned in Sections I and II, the **Triangle structure** (Fig. 2) is a causal structure $\mathcal{G}$ consisting of 3 observable variables A, B, C arranged in a triangular configuration while pair-wise sharing latent variables X, Y, Z.

Following the definition of causal compatibility from Section II, a distribution $P_{\mathcal{N}_O} = P_{ABC}$ is compatible with the Triangle structure if and only if there exists a choice of causal parameters $\left\{P_{A|X,Y}, P_{B|Y,Z}, P_{C|Z,X}, P_X, P_Y, P_Z\right\}$ such that P_{ABC} is a marginalization of P_{ABCXYZ} over X, Y, Z:

$$P_{ABC} = \sum_{X,Y,Z} P_{A|X,Y} P_{B|Y,Z} P_{C|Z,X} P_X P_Y P_Z \tag{4}$$

For certain causal structures, the conditional independence relations provided by d-separation relations are a sufficient characterization of compatibility. Recently, Henson *et al.* [16], classified the Triangle structure as an *interesting* causal structure in the sense that conditional independence relations *do not* provide a sufficient characterization of non-classicality.[3] Moreover, Wolfe *et al.* [17] concretely demonstrate that the space of classical distributions on the Triangle structure is non-convex, unlike the Bell structure. It can be argued that convexity of the Bell structure distributions is partially responsible for the wealth of knowledge that is known about the Bell structure. It is also reasonable to speculate that quantum non-classicality in the triangle can be translated directly into a computational advantage for certain computational circuits [18]. Characterizing quantum non-classicality for interesting causal structures remains an unsolved problem, and the Triangle structure is a small, non-trivial case study.

Additionally, Fritz [11] demonstrated that the Triangle structure is the *smallest* correlation scenario[4] in which their exists incompatible, quantum distributions by explicitly mentioning one which we have elected to call the Fritz distribution (see Section V). Fritz's proof of non-classicality does not involve compatibility inequalities, but instead mentions that inequalities would be helpful for characterizing non-classical in the Triangle structure. Comment (Elie Wolfe): Revisions processed up to here thusfar.

Finally Wolfe *et al.* [17] proposed the Inflation Technique (summarized in Appendix A), a novel tool for solving causal inference problems. Using the inflation technique, they were able to prove incompatibility between the Triangle structure and the W-type distribution; a quantum non-accessible distribution whose incompatibility is not witnessable by *any* entropic inequalities or any other known constraints. Therefore as a an algorithmic framework, the Inflation Technique comes closer to characterizing non-classicality in the Triangle structure that all others.

It is worth noting that a number of causal compatibility inequalities for the Triangle structure have been derived in previous works. For example, Steudel and Ay [13] derived an inequality for a series of causal structures with $n \in \mathbb{N}$ variables where at most $c \in \mathbb{N}$ share a common ancestor. The Triangle structure is a special case with $c = 2$, $n = 3$. Moreover, Henson *et al.* [16] derive a family of entropic inequalities constraining the set of compatible correlations. Most impressively, Wolfe *et al.* [17] derive a family of polynomial, 2-outcome, causal compatibility inequalities; 37 non-trivial representatives are printed in [17].

---

[3] Note that the Triangle structure possesses *zero* conditional independence relations over the observable nodes A, B, C.

[4] Correlation scenarios differ from the Bell structure in that measure setting variables are non-existent in correlation scenarios.

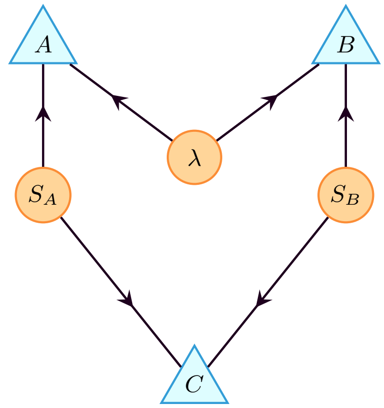


FIG. 3: The Triangle structure re-imagined to mimic the Bell structure. The measurement settings S_A, S_B are latent nodes unlike the Bell structure (Fig. 1).

As a preliminary search for quantum incompatibility in the Triangle structure, we performed numerical optimizations using bipartite qubit density matrices and 2-outcome POVM measurements against the compatibility inequalities printed in [17] as well as the entropic inequalities presented by Henson *et al.* [16]. Unfortunately, none of these inequalities could be violated. These early results suggest, although do not prove, that quantum non-classical does not exist in the Triangle structure for *two-outcome* measurements. It is therefore conjectured that all quantum non-classical distributions in the Triangle structure utilize more than 2 outcomes in at least one party. Moreover we take these observations as motivation to explore incompatibility in Triangle structure for a larger cardinalities.

In conclusion, the Triangle structure is a causal structure whose non-classicality remains an open problem. The Triangle structure is a desirable case study because there is a quantum distribution that is known to be incompatible, but no previously known inequalities were able to witness its incompatibility. This failure represents a gap in our understanding of quantum non-classicality, where recent developments (namely the Inflation Technique) have made partial progress. This document expands on our understanding of non-classicality in the Triangle structure and also our understanding on how to find inequalities using the Inflation Technique for interesting causal structures in general.

V. FRITZ DISTRIBUTION

As was originally noticed by Fritz [11], it is possible to construct quantum distributions incompatible with the Triangle structure by utilizing quantum distributions incompatible with the familiar Bell structure. To explain, imagine rearranging the triangle into the configuration depicted in Fig. 3 and contrast it with the Bell structure of Fig. 1. Evidently, under the correct relabeling, large portions of the Triangle structure resemble the Bell structure. The crucial distinction is that the measurement settings S_A, S_B are *observable variables* in the Bell structure but become *latent variables* in the Triangle structure. In order to embed distributions that are incompatible with the Bell structure into the Triangle structure, the latent nodes S_A, S_B need to be observable and independent of the shared latent node λ. This can be accomplished by having C "measure" the measurement settings S_A, S_B and announce them as an outcome. Consequently, any distribution over A, B, S_A, S_B that is incompatible with Bell structure is also incompatible with the Triangle structure provided that C is *perfectly correlated* with S_A, S_B [11]. Understandably there are numerous ways that this can be accomplished, albeit we will focus on elucidating an exemplary method.

The **Fritz distribution**, denoted P_{F}, is a quantum-accessible distribution incompatible with the Triangle structure [11]. In the Fritz distribution, each of the variables A, B, C have 4 possible outcomes. Explicitly, P_{F} can be written as:

$$\begin{aligned} P_{\mathrm{F}}(000) = P_{\mathrm{F}}(110) = P_{\mathrm{F}}(021) = P_{\mathrm{F}}(131) = P_{\mathrm{F}}(202) = P_{\mathrm{F}}(312) = P_{\mathrm{F}}(233) = P_{\mathrm{F}}(323) = \frac{1}{32}\left(2+\sqrt{2}\right) \\ P_{\mathrm{F}}(010) = P_{\mathrm{F}}(100) = P_{\mathrm{F}}(031) = P_{\mathrm{F}}(121) = P_{\mathrm{F}}(212) = P_{\mathrm{F}}(302) = P_{\mathrm{F}}(223) = P_{\mathrm{F}}(333) = \frac{1}{32}\left(2-\sqrt{2}\right) \end{aligned} \tag{5}$$

Here the notation $P_{\mathrm{F}}(abc) = P_{ABC}(abc) = P(A=a, B=b, C=c)$ is used as shorthand. The Fritz distribution is quantum-accessible in the sense that P_{F} can be implemented using a set of quantum states $\rho_{AB}, \rho_{BC}, \rho_{CA}$ and measurements M_A, M_B, M_C realized on Fig. 2 using Eq. (16) [11]. The Fritz distribution is best visualized as a $4 \times 4 \times 4$ grid of possible outcomes as depicted in Fig. 4. From this diagram, it can be seen that each of C's outcomes

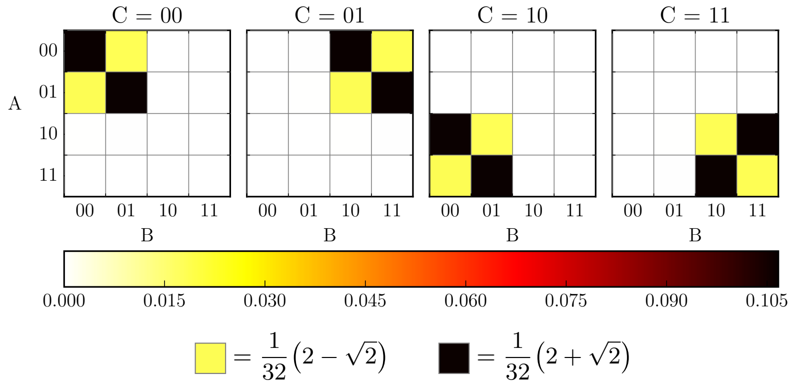


FIG. 4: The Fritz distribution visualized using a $4 \times 4 \times 4$ grid. The 4 outcomes of A, B, C are written in binary as a doublet of bits to illustrate that certain bits act as measurement pseudo-settings.

restricts the possible outcomes for A, B into a 2×2 block. If one writes the outcome labels $\{0, 1, 2, 3\}$ in binary $\{00, 01, 10, 11\}$, it can be seen that the left-hand bits for A and B (respectively denoted A_l, B_l) are fixed by the outcome of C. Pursuant to the embedding of Fig. 3 the left-hand bits emulate the measurement settings S_A and S_B and the right-hand bits emulate a two-outcome measurement performed by A, B as they would in Fig. 1. Specifically, C's bits are perfectly correlated with the left-bits of A, B; $C_l = A_l$ and $C_r = B_l$. Consequently, it is possible to define the correlation $\langle A_r B_r \rangle$ between the right-hand bits of A and B:

$$\langle A_r B_r \rangle = P_{A_r B_r}(00) + P_{A_r B_r}(11) - P_{A_r B_r}(01) - P_{A_r B_r}(10)$$

Provided that C is perfectly correlated with A_l and B_l, any Bell inequality for the Bell structure defined over A, B, S_A, S_B can be directly converted to an inequality for the Triangle structure by performing a simple relabeling: $A, B, S_A, S_B \mapsto A_r, B_r, A_l, B_l$. As an example, the famous CHSH inequality [19] (Eq. (3)) constrains the correlation between the right-bits of A and B,

$$\langle A_r B_r | C = 00 \rangle + \langle A_r B_r | C = 01 \rangle + \langle A_r B_r | C = 10 \rangle - \langle A_r B_r | C = 11 \rangle \leq 2 \tag{6}$$

Every compatible distribution P_{ABC} for which C is perfectly correlated A_l, B_l must satisfy Eq. (6). Substituting the Eq. (5) into Eq. (6) yields the maximal violation [20]: $3\left(1/\sqrt{2}\right) - \left(-1/\sqrt{2}\right) = 2\sqrt{2} \nleq 2$. Therefore, we conclude the Fritz distribution is incompatible with the Triangle structure. Before continuing it is worth noting that Eq. (5) is non-unique. Nonetheless for concreteness, Eq. (5) is taken as *the* Fritz distribution throughout this manuscript.

It is important to understand the domain in which Fritz's proof of incompatibility is valid; the details of which can found in the original work [11]. Specifically, Eq. (6) only acts as a valid causal compatibility inequalty for the Triangle structure if there exists perfect correlation between C's outcomes and the measurement pseudo-settings (left-bits) of A and B. As a result, the incompatibility proof employed by Fritz is not applicable to any distribution that significantly deviates from the Fritz distribution presented above. For example, if one combines Eq. (5) with more and more uniform noise, at what point does the resultant distribution transition from incompatibility to compatibility? This question is partially answered in Section VII A.

Upon reflection, the Fritz distribution can be considered *manufactured*. The phenomenology associated with Bell non-locality or Bell incompatibility are well understood; examining these distributions embedded in the Triangle structure offers no additional perspective onto the types of entanglement resources made accessible by quantum mechanics. Therefore, it is desirable to find incompatible quantum distributions that are qualitatively different than those previously considered for the Bell structure. Pursuant to this, Fritz [11] presented the following problem [11, Problem 2.17]:

Fritz's Problem: Find an example of non-classical quantum correlations in the Triangle structure together with a proof of its non-classicality which does not hinge on Bell's theorem.

Understandably, Fritz's Problem is stated ambiguously as it is not yet entirely clear how to distinguish between non-classical distributions on differing causal structures. The remainder of this paper focuses on our advancement toward solving Fritz's Problem. Section VIII proposes three refinements of Fritz's Problem which are to be labeled **R.1**,**R.2** and **R.3** respectively. Specifically, **R.1** embodies the degree in which the results of Section VI solve Fritz's problem while **R.2** and **R.3** characterize interpretations of Fritz's Problem that remain partially solved.

VI. TRIANGLE STRUCTURE INEQUALITIES

Equipped with the Fritz distribution (Section V) and the inflation technique (Appendix A), it is now possible to derive causal compatibility inequalities for the triangle that are violated by the Fritz distribution; the procedure for doing so is as follows. First select an inflation $\mathcal{G}'$ of the Triangle structure $\mathcal{G}$ and identify the maximal pre-injectable sets $\mathsf{PreInj}_{\mathcal{G}}(\mathcal{G}')$. These sets define the marginal scenario to be used in the Inflation Technique. Next, inflate the Fritz distribution into the pre-injectable sets as described by the Inflation lemma (Lemma 11). Finally, obtain infeasibility certificates using some linear programing software and obtain a valid causal compatibility inequality that witnesses the Fritz distribution's incompatibility.

As a first demonstration, we make use of the Wagon-Wheel inflation (Fig. 9c) to find these inequalities. The Wagon-Wheel inflation possesses 4 copies of C (C_1, C_2, C_3, C_4) and 2 copies of A (A_1, A_2) and B (B_1, B_2) arranged in the shape of a Wagon-Wheel. By inspection[5], the maximal pre-injectable sets of the Wagon-Wheel inflation are as follows:

$$\begin{array}{cc} \textbf{Maximal Pre-injectable Sets} & \textbf{Ancestral Independences} \\ \{A_2, B_1, C_3, C_1\} & \{A_2, B_1, C_3\} \perp \{C_1\} \\ \{A_1, B_1, C_4, C_2\} & \{A_1, B_1, C_4\} \perp \{C_2\} \\ \{A_1, B_2, C_1, C_3\} & \{A_1, B_2, C_1\} \perp \{C_3\} \\ \{A_2, B_2, C_2, C_4\} & \{A_2, B_2, C_2\} \perp \{C_4\} \end{array} \tag{7}$$

These maximal pre-injectable sets define a marginal scenario where $\mathcal{G}'$ is the Wagon-Wheel inflation of the Triangle structure $\mathcal{G}$.

$$\mathcal{M} = \mathsf{PreInj}_{\mathcal{G}}(\mathcal{G}') = \{\{A_2, B_1, C_3, C_1\}, \{A_1, B_1, C_4, C_2\}, \{A_1, B_2, C_1, C_3\}, \{A_2, B_2, C_2, C_4\}\} \tag{8}$$

The complete set of variables $\mathcal{J}$ is simply the set of observable nodes in the Wagon-Wheel inflation,

$$\mathcal{J} = \mathcal{N}'_O = \{A_1, A_2, B_1, B_2, C_1, C_2, C_3, C_4\}$$

Given that the Fritz distribution is defined over 4-outcomes variables, the variables in the marginal scenario are assigned 4-outcomes as well. This marginal scenario $\mathcal{M}$ then defines an incidence matrix capable of accommodating the Fritz distribution. Analogously to Theorem 4, this marginal scenario defines an incidence matrix M that has $|\mathcal{S}(\mathcal{M})| = \sum_{V\in\mathcal{M}} |\mathcal{S}(V)| = 4 \cdot 4^4 = 1024$ rows and $|\mathcal{S}(\mathcal{J})| = \prod_{v\in\mathcal{J}} |\mathcal{S}(v)| = 4^{|\mathcal{J}|} = 4^8 = 65,536$ columns. Understandably, this matrix is not reproduced here. The sheer size of the incidence matrix used here makes complete solutions of the marginal problem using tools such as linear-quantifier elimination intractable.

The marginal distribution vector $\mathcal{P}^{\mathcal{M}}$ is indexed by 1024 marginal outcomes $m \in \mathcal{S}(\mathcal{M})$ of the following *symbolic* form,

$$\mathcal{P}^{\mathcal{M}\intercal} = \underbrace{(P_{A_2B_1C_3C_1}(0000), \ldots, P_{A_1B_1C_4C_2}(3232), \ldots, P_{A_2B_2C_2C_4}(3333))}_{\text{1024 entries}} \tag{9}$$

Which can be factored into the ancestrally independent injectable sets using Eq. (7),

$$\mathcal{P}^{\mathcal{M}\intercal} = (P_{A_2B_1C_3}(000)P_{C_1}(0), \ldots, P_{A_1B_1C_4}(323)P_{C_2}(2), \ldots, P_{A_2B_2C_2}(333)P_{C_4}(3)) \tag{10}$$

And finally, each probability distribution in Eq. (10) is deflated by dropping copy-indices.

$$\mathcal{P}^{\mathcal{M}\intercal} = (P_{ABC}(000)P_C(0), \ldots, P_{ABC}(323)P_C(2), \ldots, P_{ABC}(333)P_C(3)) \tag{11}$$

This step is permitted because all of the remaining distributions in Eq. (10) are defined over the injectable sets of the Wagon-Wheel inflation. Numerically however, each of the elements of $\mathcal{P}^{\mathcal{M}}$ are replaced with numerical values corresponding to the Fritz distribution upon deflation. Pursuant to Eq. (5), $P_{ABC}(323)P_C(2)$, for example, is assigned the following numerical value.

$$P_{ABC}(323)P_C(2) = \frac{1}{32}\left(2 + \sqrt{2}\right) \cdot \frac{1}{4} \simeq 0.0267$$

[5] Computing the injectable and pre-injectable sets computationally is possible and a technique is discussed in Wolfe *et al.* [17].

The same is applied to all other entries of $\mathcal{P}^{\mathcal{M}}$. Finally, this numerical version of $\mathcal{P}^{\mathcal{M}}$ and M are subjected to linear programing software and an infeasibility certificate y was obtained. Below is an example of an infeasibility inequality obtained over the entries of Eq. (11).

$$\begin{aligned}
2P_{ABC}(203)P_C(0) + 2P_{ABC}(303)P_C(0) + P_{ABC}(003)P_C(0) + P_{ABC}(013)P_C(0)+\\
P_{ABC}(020)P_C(2) + P_{ABC}(022)P_C(2) + P_{ABC}(023)P_C(0) + P_{ABC}(023)P_C(2)+\\
P_{ABC}(030)P_C(2) + P_{ABC}(031)P_C(2) + P_{ABC}(032)P_C(2) + P_{ABC}(033)P_C(0)+\\
P_{ABC}(033)P_C(2) + P_{ABC}(103)P_C(0) + P_{ABC}(113)P_C(0) + P_{ABC}(123)P_C(0)+\\
P_{ABC}(133)P_C(0) + P_{ABC}(200)P_C(0) + P_{ABC}(200)P_C(1) + P_{ABC}(200)P_C(2)+\\
P_{ABC}(200)P_C(3) + P_{ABC}(201)P_C(0) + P_{ABC}(201)P_C(1) + P_{ABC}(201)P_C(2)+\\
P_{ABC}(201)P_C(3) + P_{ABC}(203)P_C(1) + P_{ABC}(203)P_C(2) + P_{ABC}(203)P_C(3)+\\
P_{ABC}(213)P_C(0) + P_{ABC}(223)P_C(0) + P_{ABC}(300)P_C(0) + P_{ABC}(300)P_C(1)+\\
P_{ABC}(300)P_C(2) + P_{ABC}(300)P_C(3) + P_{ABC}(301)P_C(0) + P_{ABC}(301)P_C(1)+\\
P_{ABC}(301)P_C(2) + P_{ABC}(301)P_C(3) + P_{ABC}(302)P_C(1) + P_{ABC}(303)P_C(1)+\\
P_{ABC}(303)P_C(2) + P_{ABC}(303)P_C(3) + P_{ABC}(313)P_C(0) - P_{ABC}(000)P_C(3) \geq 0
\end{aligned} \tag{12}$$

By construction, every marginal model $\mathcal{P}^{\mathcal{M}} = \{P_{ABC}, P_{AB}, P_{BC}, P_{AC}, P_A, P_B, P_C\}$ compatible with the Triangle structure must satisfy Eq. (12). The Fritz distribution however violates Eq. (12).

Notice the final term $P_{ABC}(000)P_C(3)$ has a coefficient -1. For the Fritz distribution, $P_{ABC}(000)P_C(3) = \frac{1}{128}\left(2+\sqrt{2}\right) \simeq 0.0267$ but all of the remaining terms sum to $\simeq 0.0136$ meaning that the Fritz distribution violates this inequality with violation $-0.0129\ldots \not\geq 0$.

It is possible to simplify this inequality by marginalizing over sets of terms such as $P_{ABC}(300)P_C(0)+P_{ABC}(300)P_C(1)+P_{ABC}(300)P_C(2)+P_{ABC}(300)P_C(3) = P_{ABC}(300)$. Upon doing so, one obtains the following,

$$\begin{aligned}
P_{ABC}(000)P_C(3) + P_{ABC}(021)P_C(2) + P_{ABC}(202) + P_{ABC}(302) + P_{ABC}(233)P_C(0) + P_{AC}(33)P_C(0) \leq\\
P_{ABC}(303)P_C(0) + P_{ABC}(313)P_C(0) + P_{ABC}(302)P_C(1)+\\
+P_{AB}(02)P_C(2) + P_{AB}(03)P_C(2) + P_{AB}(20) + P_{AB}(30) + P_C(3)P_C(0)
\end{aligned}$$

There are a couple of important concepts to understand. First, the Fritz distribution is known to be incompatible with the Triangle structure due to the perfect correlation between C and the measurement pseudo-settings. The inequality contained in Eq. (12) represents a significant advancement in our understanding of non-classicality in the Triangle structure. For example, Eq. (12) is violated by distributions that do not exhibit perfect correlation between C and A_l, B_l. Moreover, although the existence of an inequality capable of witnessing Fritz distribution is guaranteed, there was no guarantee that the marginal model associated with the Wagon-Wheel inflation (Eq. (8)) would be capable of producing such an inequality. In fact, smaller inflations of the Triangle structure such as the Spiral inflation (Fig. 9b) and Cut inflation (Fig. 9a) are *not* able to produce causal compatibility inequalities that witness the Fritz distribution.

Larger inflations such as the Web inflation (Fig. 9d) were also used to generate causal compatibility inequalities that are violated by the Fritz distribution. The set of maximal pre-injectable sets for the Web inflation is considerably larger than Eq. (7).

$$\begin{array}{cc}
\textbf{Maximal Pre-injectable Sets} & \textbf{Ancestral Independences} \\
\{A_1, B_1, C_1, A_4, B_4, C_4\} & \{A_1, B_1, C_1\} \perp \{A_4, B_4, C_4\} \\
\{A_1, B_2, C_3, A_4, B_3, C_2\} & \{A_1, B_2, C_3\} \perp \{A_4, B_3, C_2\} \\
\{A_2, B_3, C_1, A_3, B_2, C_4\} & \{A_2, B_3, C_1\} \perp \{A_3, B_2, C_4\} \\
\{A_2, B_4, C_3, A_3, B_1, C_2\} & \{A_2, B_4, C_3\} \perp \{A_3, B_1, C_2\} \\
\{A_1, B_3, C_4\} & \{A_1\} \perp \{B_3\} \perp \{C_4\} \\
\{A_1, B_4, C_2\} & \{A_1\} \perp \{B_4\} \perp \{C_2\} \\
\{A_2, B_1, C_4\} & \{A_2\} \perp \{B_1\} \perp \{C_4\} \\
\{A_2, B_2, C_2\} & \{A_2\} \perp \{B_2\} \perp \{C_2\} \\
\{A_3, B_3, C_3\} & \{A_3\} \perp \{B_3\} \perp \{C_3\} \\
\{A_3, B_4, C_1\} & \{A_3\} \perp \{B_4\} \perp \{C_1\} \\
\{A_4, B_1, C_3\} & \{A_4\} \perp \{B_1\} \perp \{C_3\} \\
\{A_4, B_2, C_1\} & \{A_4\} \perp \{B_2\} \perp \{C_1\}
\end{array} \tag{13}$$

Again, analogously to Theorem 4, this marginal scenario defines an incidence matrix M that has $|\mathcal{S}(\mathcal{M})| = 4{\cdot}4^6{+}8{\cdot}4^3 = 16,896$ rows, $|\mathcal{S}(\mathcal{J})| = 4^{12} = 16,777,216$ and $201,326,592$ non-zero entires. Using this incidence matrix, a number of incompatibility certificates were derived. For inspection, one is reproduced in Eq. (14).

In summary, these causal compatibility inequalities represent a completion of objective Item **R.1** from Section V. These inequalities are derived without making use of Bell's theorem and instead use the Inflation Technique which is completely independent of Bell's theorems. Additionally, the derivation of causal compatibility inequalities allows one prove incompatibility even when perfect correlations are absent. Moreover, they allow one to probe the amount of noise the Fritz distribution can sustain before becoming compatible (see Section VII). Unfortunately, infeasibility certificates are found by linear programing software to certify the infeasibility of a *particular distribution.* This means that inequalities like Eq. (12), the one presented in Eq. (14), and others are specifically manufactured in order witness a particular type of quantum non-classicality. In an effort to derive inequalities that could lead to potentially new forms of non-classicality, we searched for inequalities that did not admit biased towards the Fritz distribution.

The following is a causal compatibility inequality for the Triangle structure that is violated by the Fritz distribution and was derived using the Web inflation of Fig. 9d.

$$
\begin{gathered}
P(000)P(202) + P(000)P(212) + P(202)P(233) + P(302)P(312) \\
\leq \\
2P(123)P(210) + 2P(123)P(310) + 2P(130)P(213) + 2P(133)P(210) + 2P(133)P(310) + P(000)P(003) + P(000)P(013) + P(000)P(023)+ \\
P(000)P(033) + P(000)P(103) + P(000)P(113) + P(000)P(123) + P(000)P(133) + P(000)P(203) + P(000)P(213) + P(003)P(010)+ \\
P(003)P(020) + P(003)P(030) + P(003)P(100) + P(003)P(110) + P(003)P(120) + P(003)P(130) + P(003)P(200) + P(003)P(210)+ \\
P(003)P(220) + P(003)P(230) + P(003)P(300) + P(003)P(310) + P(003)P(320) + P(003)P(330) + P(010)P(013) + P(010)P(021)+ \\
P(010)P(023) + P(010)P(033) + P(010)P(100) + P(010)P(103) + P(010)P(113) + P(010)P(123) + P(010)P(133) + P(010)P(203)+ \\
P(010)P(213) + P(013)P(020) + P(013)P(030) + P(013)P(100) + P(013)P(110) + P(013)P(120) + P(013)P(130) + P(013)P(200)+ \\
P(013)P(210) + P(013)P(220) + P(013)P(230) + P(013)P(300) + P(013)P(310) + P(013)P(320) + P(013)P(330) + P(020)P(023)+ \\
P(020)P(033) + P(020)P(103) + P(020)P(113) + P(020)P(123) + P(020)P(133) + P(020)P(203) + P(020)P(210) + P(020)P(211)+ \\
P(020)P(212) + P(020)P(223) + P(020)P(233) + P(020)P(310) + P(020)P(311) + P(020)P(312) + P(020)P(313) + P(021)P(100)+ \\
P(021)P(121) + P(021)P(210) + P(021)P(211) + P(021)P(213) + P(021)P(310) + P(021)P(311) + P(021)P(313) + P(022)P(210)+ \\
P(022)P(211) + P(022)P(212) + P(022)P(213) + P(022)P(310) + P(022)P(311) + P(022)P(312) + P(022)P(313) + P(023)P(030)+ \\
P(023)P(100) + P(023)P(110) + P(023)P(120) + P(023)P(130) + P(023)P(200) + P(023)P(211) + P(023)P(212) + P(023)P(213)+ \\
P(023)P(220) + P(023)P(230) + P(023)P(300) + P(023)P(311) + P(023)P(312) + P(023)P(313) + P(023)P(320) + P(023)P(330)+ \\
P(030)P(033) + P(030)P(103) + P(030)P(113) + P(030)P(123) + P(030)P(133) + P(030)P(203) + P(030)P(210) + P(030)P(211)+ \\
P(030)P(212) + P(030)P(223) + P(030)P(233) + P(030)P(310) + P(030)P(311) + P(030)P(312) + P(030)P(313) + P(031)P(210)+ \\
P(031)P(211) + P(031)P(213) + P(031)P(223) + P(031)P(310) + P(031)P(311) + P(031)P(313) + P(032)P(210) + P(032)P(211)+ \\
P(032)P(212) + P(032)P(213) + P(032)P(310) + P(032)P(311) + P(032)P(312) + P(032)P(313) + P(033)P(100) + P(033)P(110)+ \\
P(033)P(120) + P(033)P(130) + P(033)P(200) + P(033)P(211) + P(033)P(212) + P(033)P(213) + P(033)P(220) + P(033)P(230)+ \\
P(033)P(300) + P(033)P(311) + P(033)P(312) + P(033)P(313) + P(033)P(320) + P(033)P(330) + P(100)P(103) + P(100)P(113)+ \\
P(100)P(123) + P(100)P(133) + P(100)P(203) + P(100)P(213) + P(103)P(110) + P(103)P(120) + P(103)P(130) + P(103)P(200)+ \\
P(103)P(210) + P(103)P(220) + P(103)P(230) + P(103)P(300) + P(103)P(310) + P(103)P(320) + P(103)P(330) + P(110)P(113)+ \\
P(110)P(123) + P(110)P(133) + P(110)P(203) + P(110)P(213) + P(113)P(120) + P(113)P(130) + P(113)P(200) + P(113)P(210)+ \\
P(113)P(220) + P(113)P(230) + P(113)P(300) + P(113)P(310) + P(113)P(320) + P(113)P(330) + P(120)P(123) + P(120)P(133)+ \\
P(120)P(203) + P(120)P(210) + P(120)P(211) + P(120)P(212) + P(120)P(223) + P(120)P(233) + P(120)P(310) + P(120)P(311)+ \\
P(120)P(312) + P(120)P(313) + P(121)P(131) + P(121)P(210) + P(121)P(211) + P(121)P(213) + P(121)P(310) + P(121)P(311)+ \\
P(121)P(312) + P(121)P(313) + P(122)P(210) + P(122)P(211) + P(122)P(212) + P(122)P(213) + P(122)P(310) + P(122)P(311)+ \\
P(122)P(312) + P(122)P(313) + P(123)P(130) + P(123)P(200) + P(123)P(211) + P(123)P(212) + P(123)P(213) + P(123)P(220)+ \\
P(123)P(230) + P(123)P(300) + P(123)P(311) + P(123)P(312) + P(123)P(313) + P(123)P(320) + P(123)P(330) + P(130)P(133)+ \\
P(130)P(203) + P(130)P(210) + P(130)P(211) + P(130)P(212) + P(130)P(223) + P(130)P(233) + P(130)P(310) + P(130)P(311)+ \\
P(130)P(312) + P(130)P(313) + P(131)P(210) + P(131)P(211) + P(131)P(213) + P(131)P(310) + P(131)P(311) + P(131)P(313)+ \\
P(132)P(210) + P(132)P(211) + P(132)P(212) + P(132)P(213) + P(132)P(310) + P(132)P(311) + P(132)P(312) + P(132)P(313)+ \\
P(133)P(200) + P(133)P(211) + P(133)P(212) + P(133)P(213) + P(133)P(220) + P(133)P(230) + P(133)P(300) + P(133)P(311)+ \\
P(133)P(312) + P(133)P(313) + P(133)P(320) + P(133)P(330) + P(200)P(203) + P(200)P(213) + P(200)P(223) + P(200)P(233)+ \\
P(200)P(303) + P(200)P(313) + P(200)P(323) + P(200)P(333) + P(203)P(210) + P(203)P(220) + P(203)P(230) + P(203)P(300)+ \\
P(203)P(310) + P(203)P(320) + P(203)P(330) + P(210)P(213) + P(210)P(223) + P(210)P(233) + P(210)P(303) + P(210)P(313)+ \\
P(210)P(323) + P(210)P(333) + P(212)P(212) + P(212)P(333) + P(213)P(220) + P(213)P(230) + P(213)P(300) + P(213)P(310)+ \\
P(213)P(320) + P(213)P(330) + P(220)P(223) + P(220)P(233) + P(220)P(303) + P(220)P(313) + P(220)P(323) + P(220)P(333)+ \\
P(223)P(230) + P(223)P(300) + P(223)P(310) + P(223)P(320) + P(223)P(330) + P(230)P(233) + P(230)P(303) + P(230)P(313)+ \\
P(230)P(323) + P(230)P(333) + P(233)P(300) + P(233)P(310) + P(233)P(320) + P(233)P(330) + P(233)P(333) + P(300)P(303)+ \\
P(300)P(313) + P(300)P(323) + P(300)P(333) + P(303)P(310) + P(303)P(320) + P(303)P(330) + P(310)P(313) + P(310)P(323)+ \\
P(310)P(333) + P(313)P(320) + P(313)P(330) + P(320)P(323) + P(320)P(333) + P(323)P(330) + P(330)P(333)
\end{gathered}
\tag{14}
$$

Note: $P(abc)$ is shorthand for $P_{ABC}(abc)$. The following is a symmetric causal compatibility inequality for the Triangle structure that is violated by the Fritz distribution and was derived using the Web inflation of Fig. 9d.

$$
\begin{gathered}
2[P(312)P(312)]_6+3[P(000)P(323)]_3+3[P(131)P(323)]_3+[P(000)P(121)]_3+[P(110)P(333)]_3+[P(202)P(202)]_3 \\
\leq \\
12[P(002)P(031)]_6+12[P(010)P(330)]_6+12[P(032)P(210)]_6+12[P(101)P(331)]_6+12[P(122)P(321)]_6+12[P(130)P(312)]_6+ \\
12[P(202)P(220)]_6+16[P(002)P(120)]_6+16[P(002)P(200)]_6+16[P(002)P(220)]_3+16[P(002)P(230)]_6+16[P(002)P(302)]_6+ \\
16[P(002)P(303)]_6+16[P(002)P(330)]_3+16[P(003)P(032)]_6+16[P(003)P(033)]_6+16[P(003)P(231)]_6+16[P(003)P(300)]_6+ \\
16[P(003)P(310)]_6+16[P(003)P(320)]_6+16[P(003)P(322)]_6+16[P(012)P(202)]_6+16[P(012)P(210)]_6+16[P(012)P(313)]_6+ \\
16[P(013)P(120)]_6+16[P(013)P(130)]_6+16[P(013)P(203)]_6+16[P(013)P(220)]_6+16[P(020)P(101)]_3+16[P(020)P(103)]_6+ \\
16[P(020)P(312)]_6+16[P(021)P(311)]_6+16[P(021)P(332)]_6+16[P(022)P(203)]_6+16[P(022)P(302)]_6+16[P(022)P(303)]_6+ \\
16[P(022)P(311)]_3+16[P(022)P(313)]_6+16[P(023)P(103)]_6+16[P(023)P(222)]_6+16[P(023)P(232)]_6+16[P(030)P(201)]_6+ \\
16[P(031)P(301)]_6+16[P(031)P(332)]_6+16[P(032)P(311)]_6+16[P(033)P(202)]_6+16[P(033)P(300)]_3+16[P(033)P(312)]_6+ \\
16[P(033)P(331)]_6+16[P(033)P(333)]_3+16[P(102)P(130)]_6+16[P(102)P(220)]_6+16[P(102)P(230)]_6+16[P(102)P(310)]_6+ \\
16[P(103)P(110)]_6+16[P(103)P(133)]_6+16[P(113)P(130)]_6+16[P(113)P(220)]_3+16[P(113)P(321)]_6+16[P(120)P(201)]_6+ \\
16[P(120)P(223)]_6+16[P(120)P(303)]_6+16[P(120)P(313)]_6+16[P(123)P(230)]_6+16[P(123)P(302)]_6+16[P(123)P(330)]_6+ \\
16[P(130)P(203)]_6+16[P(130)P(223)]_6+16[P(131)P(303)]_3+16[P(132)P(202)]_6+16[P(133)P(202)]_6+16[P(201)P(330)]_6+ \\
16[P(203)P(330)]_6+16[P(210)P(213)]_6+16[P(222)P(303)]_3+16[P(223)P(310)]_6+16[P(231)P(301)]_6+16[P(232)P(330)]_6+ \\
16[P(233)P(303)]_6+16[P(300)P(331)]_6+16[P(301)P(330)]_6+16[P(302)P(331)]_6+16[P(313)P(320)]_6+20[P(132)P(301)]_6+ \\
20[P(133)P(301)]_6+24[P(033)P(310)]_6+24[P(102)P(231)]_6+24[P(103)P(331)]_6+24[P(130)P(303)]_6+2[P(010)P(031)]_6+ \\
2[P(100)P(100)]_3+2[P(100)P(223)]_6+2[P(110)P(223)]_3+2[P(121)P(312)]_6+2[P(212)P(302)]_6+3[P(000)P(333)]_1+ \\
3[P(031)P(312)]_6+3[P(223)P(333)]_3+3[P(302)P(312)]_6+4[P(000)P(022)]_3+4[P(000)P(212)]_3+4[P(000)P(301)]_6+ \\
4[P(001)P(031)]_6+4[P(003)P(211)]_6+4[P(010)P(132)]_6+4[P(010)P(311)]_6+4[P(011)P(211)]_3+4[P(012)P(021)]_6+ \\
4[P(022)P(122)]_3+4[P(022)P(222)]_3+4[P(022)P(232)]_6+4[P(031)P(233)]_6+4[P(100)P(122)]_3+4[P(100)P(230)]_6+ \\
4[P(101)P(131)]_3+4[P(110)P(123)]_6+4[P(110)P(320)]_6+4[P(111)P(311)]_3+4[P(111)P(321)]_6+4[P(122)P(122)]_3+ \\
4[P(122)P(220)]_6+4[P(122)P(232)]_6+4[P(123)P(131)]_6+4[P(132)P(332)]_6+4[P(201)P(201)]_6+4[P(201)P(311)]_6+ \\
4[P(202)P(311)]_6+4[P(202)P(321)]_6+4[P(212)P(310)]_6+4[P(212)P(331)]_6+4[P(223)P(312)]_6+4[P(302)P(333)]_6+ \\
4[P(311)P(323)]_6+5[P(010)P(333)]_3+5[P(031)P(100)]_6+5[P(223)P(223)]_3+5[P(233)P(333)]_3+8[P(000)P(030)]_3+ \\
8[P(000)P(032)]_6+8[P(003)P(111)]_3+8[P(003)P(331)]_3+8[P(012)P(030)]_6+8[P(012)P(110)]_6+8[P(012)P(221)]_6+ \\
8[P(013)P(021)]_6+8[P(013)P(022)]_6+8[P(013)P(311)]_6+8[P(020)P(100)]_6+8[P(020)P(110)]_6+8[P(020)P(112)]_6+ \\
8[P(020)P(210)]_6+8[P(022)P(121)]_6+8[P(022)P(130)]_6+8[P(023)P(112)]_6+8[P(023)P(213)]_6+8[P(030)P(210)]_6+ \\
8[P(030)P(311)]_6+8[P(032)P(111)]_6+8[P(033)P(121)]_6+8[P(033)P(131)]_6+8[P(100)P(132)]_6+8[P(102)P(321)]_6+ \\
8[P(110)P(300)]_6+8[P(111)P(130)]_6+8[P(112)P(123)]_6+8[P(112)P(331)]_3+8[P(113)P(200)]_6+8[P(113)P(331)]_3+ \\
8[P(120)P(220)]_6+8[P(120)P(302)]_6+8[P(120)P(310)]_6+8[P(121)P(211)]_6+8[P(121)P(301)]_6+8[P(121)P(303)]_3+ \\
8[P(121)P(320)]_6+8[P(122)P(201)]_6+8[P(122)P(203)]_6+8[P(123)P(200)]_6+8[P(123)P(203)]_6+8[P(123)P(222)]_6+ \\
8[P(123)P(313)]_6+8[P(130)P(212)]_6+8[P(131)P(310)]_6+8[P(132)P(310)]_6+8[P(132)P(312)]_6+8[P(132)P(322)]_6+ \\
8[P(133)P(200)]_3+8[P(133)P(213)]_6+8[P(133)P(222)]_3+8[P(133)P(300)]_3+8[P(133)P(312)]_6+8[P(133)P(322)]_3+ \\
8[P(201)P(313)]_6+8[P(210)P(320)]_6+8[P(211)P(221)]_6+8[P(211)P(320)]_6+8[P(213)P(322)]_6+8[P(220)P(223)]_3+ \\
8[P(221)P(312)]_6+8[P(222)P(312)]_6+8[P(231)P(312)]_6+8[P(232)P(313)]_3+8[P(233)P(313)]_6+8[P(302)P(332)]_6+ \\
8[P(311)P(321)]_6+[P(000)P(121)]_3+[P(000)P(323)]_3+[P(110)P(333)]_3+[P(121)P(333)]_3+[P(131)P(323)]_3+ \\
[P(312)P(312)]_6
\end{gathered}
\tag{15}
$$

Note: $P(abc)$ is shorthand for $P_{ABC}(abc)$ and $[P(113)P(330)]_3$ is shorthand for the sum over permutations:

$$[P(113)P(330)]_3 \equiv P(113)P(330)+P(131)P(303)+P(311)P(033)$$

VII. NUMERICAL OPTIMIZATIONS & NOISY DISTRIBUTIONS

Causal compatibility inequalities for the Triangle structure (such as those derived in Appendix A 3) offer an avenue into studying quantum non-classicality. In this section, we make use of two methods, exposure to noise and numerical optimization, to further understand the nature of quantum non-classicality in connection to compatibility inequalities. In order to compare different inequalities simultaneously, our convention is to rearrange each inequality I so $I(P) < 0$ represent a violation of I while $I(P) \leq 0$ represents satisfaction of I (with $I(P) = 0$ corresponding to complete saturation of the inequality). For the subsequent analysis, 4 inequalities where chosen. First, I_0 is the inequality

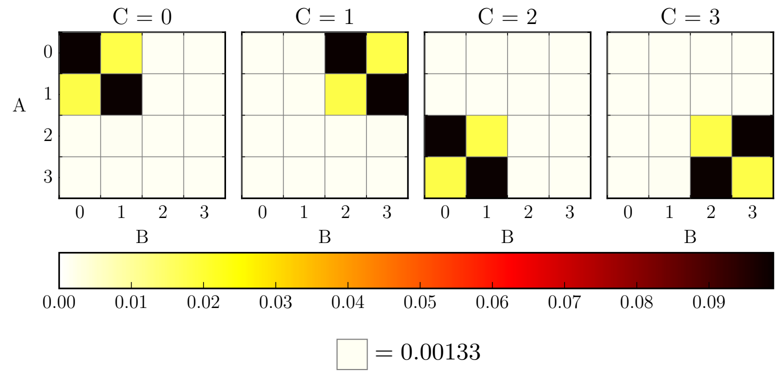


FIG. 5: $\mathcal{N}_{0.085}$: A noisy yet still non-classical variant of the Fritz distribution.

derived using the Wagon-Wheel wheel inflation Eq. (12). The other three inequalities I_1, I_2, I_3 were all derived using the Web inflation; I_1 can be found in Eq. (14) while I_3 is the symmetric inequality in Eq. (15). Inequality I_2 is not reproduced anywhere in this document but is similar to Eq. (14).

A. Noisy Distributions

In his original work, Fritz [11] was able to find quantum non-classicality in the Triangle structure by embedding Bell non-classicality. As was mentioned in Section V, this embedding required a very idealistic condition to hold: perfect correlations between C and A_l, B_l. From an experimental point of view, it is impossible to use Fritz's original argument to confirm non-classicality because every observable distribution is subject to noise. It is possible to minimize noise by developing more accurate measurement channels but perfect correlations are too idealistic. Causal compatibility inequalities not only permit there to be noise within a set of observations, but they also provide a means of quantifying the amount of noise that is allowed before non-classicality breaks down.

For Triangle structure, we define a completely noisy observation as the uniform probability distribution $\mathcal{U}(abc) = \mathcal{U}_{ABC}(abc) = \frac{1}{64} \quad \forall a, b, c \in \{0, 1, 2, 3\}$. Since the uniform distribution $\mathcal{U}$ is trivially compatible with the Triangle structure, the objective of this section is to determine how much uniform noise needs to be added to the Fritz distribution P_{F} before it transitions from incompatible to compatible. We do this by defining an affine noise parameter ε such that $0 \leq \varepsilon \leq 1$ along with the **noisy variant** $\mathcal{N}_\varepsilon$ of the Fritz distribution,

$$\mathcal{N}_\varepsilon = (1 - \varepsilon) P_{\mathrm{F}} + \varepsilon \mathcal{U}$$

Therefore as ε varies from 0 to 1, more and more noise is added to the Fritz distribution and $\mathcal{N}_\varepsilon$ transitions from an incompatible distribution $\mathcal{N}_0 = P_{\mathrm{F}}$ to a compatible distribution $\mathcal{N}_1 = \mathcal{U}$. For each inequality I and distribution P, let $I(P)$ denote a scalar value representing the amount of violation seen by P. Our tests suggest that inequality I_1 (Eq. (14)) is most robust to noise and demonstrates that the Fritz distribution remains incompatible with the Triangle structure up to at least a noise parameter of $\varepsilon \simeq 0.085$. The associated distribution $\mathcal{N}_{0.085}$ is plotted in Fig. 5. Of course, there remains the possibility that another inequality will be able to withstand a larger degree of noise than I_1.

B. Numerical Optimizations

Causal compatibility inequalities act as a *measure* of non-classicality. If two non-classical distributions P_1 and P_2 violate the same inequality I with violations $I(P_1)$ and $I(P_2)$ respectively, it is somewhat reasonable to say that P_1 is *more* non-classical than P_2 whenever $I(P_1) < I(P_2)$. However, this statement is accompanied by a few caveats. First, it is possible (and frequently the case) for $I(P_1) < I(P_2)$ to hold for one inequality I but have $\tilde{I}(P_2) < \tilde{I}(P_1)$ hold for another inequality $\tilde{I}$. In this case, all one can claim is that P_1 is more incompatible than P_2 with respective to I but the contrary is true with respect to $\tilde{I}$. Second, not all inequalities I are *good* measures of non-classicality. In the case of the Bell structure, the linear inequality constraints associated with the facets of the marginal polytope are the best measures of non-classicality because the facets form a necessary and sufficient of constraints with all other inequalities following from positive mixtures. However, good measures of incompatibility are less understood in the

Triangle structure where the space of classical correlations is non-convex. At the very least, the inequalities derived in this report are likely *not* the best measures of incompatibility simply because they have be curated toward the Fritz distribution. Nonetheless, these inequalities act the *first* measures of non-classicality in the Triangle structure.

In order to find quantum distributions that are more non-classical than the Fritz, we performed numerical optimizations against each inequality I by parameterizing the space of quantum-accessible probability distributions that can be realized on the Triangle Structure Fig. 2 and thus expressed as follows:

$$P_{ABC}(abc) = \mathrm{Tr}[\Pi^{\intercal} \rho_{AB} \otimes \rho_{BC} \otimes \rho_{CA} \Pi M_{A,a} \otimes M_{B,b} \otimes M_{C,c}] \tag{16}$$

Where $\rho_{AB}, \rho_{BC}, \rho_{CA}$ are bipartite density matrices, M_A, M_B, M_C are generic measurements sets and Π is a permutation matrix that aligns the states and measurements appropriately. In order to parameterize all such distributions, we elect to parameterize the states and measurements separately. In order to qualify the scope of Eq. (16) and associated computational complexity of the parameterization, there are a two restrictions that are made with justification. Motivated by the fact that the Fritz distribution (Section V) only requires qubit states, the states ρ are taken to be bipartite *qubit* states. Qubit states are more computationally feasible compared to n-dimensional states whereby the joint density matrix $\rho_{AB} \otimes \rho_{BC} \otimes \rho_{CA}$ becomes an $n^6 \times n^6$ matrix. Additionally, we restrict our focus to projective-valued measures (PVMs) instead of projective-operator valued measures (POVMs) for three reasons. First, Section V demonstrates that PVMs are sufficient for witnessing incompatible quantum distributions in the Triangle structure. Second, although generating k-outcome POVM measurements is possible using rejection sampling techniques [21], a valid, unbiased parameterization was not found for $k > 2$. Finally, PVMs provide considerable computational advantage over POVMs as they permit Eq. (16) to be re-written as Eq. (17).

$$P_{ABC}(abc) = \langle m_{A,a} m_{B,b} m_{C,c} | \Pi^{\intercal} \rho_{AB} \otimes \rho_{BC} \otimes \rho_{CA} \Pi | m_{A,a} m_{B,b} m_{C,c} \rangle \tag{17}$$

Although there are numerous techniques that can used when parameterizing quantum states and measurements [4, 21–25], a single technique by Spengler *et al.* [26] was found to be most computationally suitable for our purposes. Spengler *et al.* [26] demonstrated that all $d \times d$ unitary matrices U can be parameterized without degeneracy as follows:

$$U = \left[\prod_{m=1}^{d-1} \left(\prod_{n=m+1}^{d} \exp(iP_n \lambda_{n,m}) \exp(i\sigma_{m,n} \lambda_{m,n}) \right) \right] \cdot \left[\prod_{l=1}^{d} \exp(iP_l \lambda_{l,l}) \right] \tag{18}$$

Where the real valued parameters $\lambda = \{\lambda_{n,m} \mid n, m \in 1, \ldots, d\}$ have periodicities $\lambda_{m,n} \in \left[0, \frac{\pi}{2}\right]$ for $m < n$ and $\lambda_{m,n} \in [0, 2\pi]$ for $m \geq n$. Moreover, P_l are one-dimensional projective operators $P_l = |l\rangle\langle l|$ and the $\sigma_{m,n}$ are generalized anti-symmetric σ-matrices $\sigma_{m,n} = -i|m\rangle\langle n| + i|n\rangle\langle m|$ where $1 \leq m < n \leq d$. This parameterization has the useful feature that each of the real-valued parameters $\lambda_{n,m}$ a direct and intuitive physical affect on each element of a computational basis $\{|1\rangle, \ldots, |d\rangle\}$. Explicitly, $\exp(i\sigma_{m,n}\lambda_{m,n})$ applies a rotation to the sub-space spanned by $|m\rangle$ and $|n\rangle$ for $m < n$. Analogously, $\exp(iP_n\lambda_{n,m})$ generates the relative phase between $|m\rangle$ and $|n\rangle$ for $m > n$ and $\exp(iP_l\lambda_{l,l})$ fixes the global phase of $|l\rangle$. Finally, although not explicitly mentioned in [26], it is possible to remove the reliance on the computationally expensive matrix exponential operations in Eq. (18) [27].

By parameterizing unitary matrices, it becomes possible to parameterize d-dimensional density matrices and d-element PVMs by recognizing that any orthonormal basis $\{|\psi_j\rangle\}$ (where $1 \leq j \leq d$) can be transformed into the computational basis $\{|j\rangle\}$ by a unitary transformation U, i.e. $U|\psi_j\rangle = |j\rangle$. First consider a d-element PVM $M = \{|m_j\rangle\langle m_j| \mid 1 \leq j \leq d\}$. Since $\{|m_j\rangle\}$ forms an orthonormal basis, one can parameterize M by writing $M = \left\{U^\dagger |j\rangle\langle j| U \mid 1 \leq j \leq d\right\}$ and parameterizing U using Eq. (18). This method was inspired by the measurement seeding method for iterative optimization used by Pál and Vértesi [28]. Analogously, this argument can be extended to full-rank d-dimensional density matrices ρ by performing a spectral decomposition $\rho = \sum_{j=1}^{d} p_j |p_j\rangle\langle p_j|$ into eigenvalues $\{p_j\}$ and eigenstates $\{|p_j\rangle\}$. Since $\mathrm{Tr}(\rho) = \sum_{j=1}^{d} p_j = 1$ the eigenvalues of ρ are parameterized without degeneracy using a tuple of $d-1$ real-valued parameters with periodicity $[0, 2\pi]$ using hyper-spherical coordinates [22, 26]. Additionally, since ρ is Hermitian, the eigenstates $\{|p_j\rangle\}$ form an orthonormal basis and therefore the eigenstates are analogously parameterized using Eq. (18): $\rho = \sum_{j=1}^{d} p_j U^\dagger |j\rangle\langle j| U$. For our purposes, we have set $d = 4$ and fixed $\lambda_{l,l} = 0$ for $1 \leq l \leq d$ in Eq. (18) because the global phase contributions are irrelevant for Eq. (17).

To quantify the degree in which each inequality I could be further violated, we define a relative violation V_F to be the maximum violation relative to the violation achieved by the Fritz distribution, i.e. $V_F \equiv \min_P\{I(P)\}/I(P_F)$. The first inequality in consideration is I_1; the non-symmetric Web inflation inequality presented in Eq. (14). At best, numerical optimizations converged to a distribution (Fig. 6) with relative violation $V_F \simeq 1.501$. Almost immediately, it is evident that Fig. 6 closely resembles the Fritz distribution. In fact, Fig. 6 and Fig. 4 share the exact same *possibilistic structure*, i.e. the same set of possible events. This result is not entirely unexpected. As was mentioned

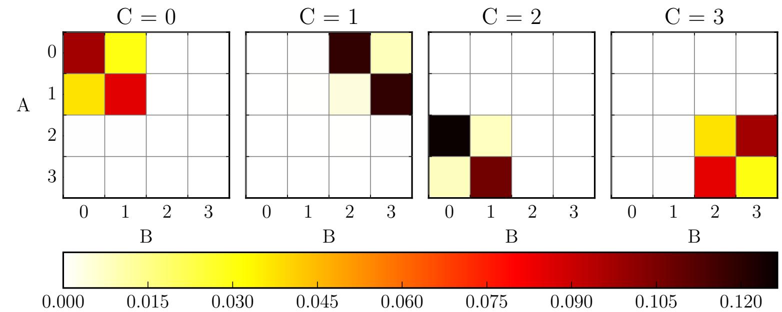


FIG. 6: Probability distribution that maximizes violation of I_1 from Eq. (14) ($V_F \simeq 1.501$).

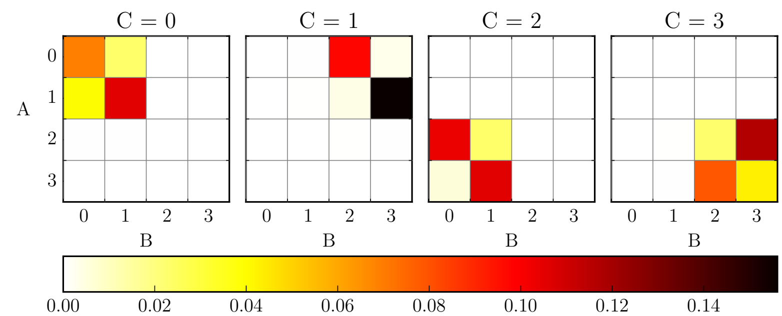


FIG. 7: Probability distribution that maximizes violation of I_3 from Eq. (15) ($V_F \simeq 2.61$).

in Section VI, these inequalities are derived specifically to prove the incompatibility of the Fritz distribution and the Triangle structure. Consequently, it seems that I_1 only has the capacity to demonstrate incompatibility for distributions that resemble the Fritz distribution very closely.

An alternative explanation for qualitative similarities between Fig. 6 and Fig. 4 is related to the sensitively of the initial parameters to our optimization. Above it was shown that our parameter space of choice was 2π-periodic in all entries. Somewhat surprisingly, when supplied with randomly sampled initial parameters, all optimization methods consistently converged to inequality *saturation* instead of inequality *violation* which is achievable by the Fritz distribution P_{F}. Instead, it was observed that only parameters close enough to the *Fritz parameters* (i.e. any non-unique set of parameters that induce the Fritz distribution) were able to achieve violation. Since the inequalities themselves are polynomial and the space of quantum-accessible distributions is non-convex, a number of optimization methods were used in an attempt to eliminate any issues associated with local minima or ill-conditioned convergence. Specifically, the BFGS method [29, p.142] and the Nelder-Mead simplex method [29, p.238] were used along with a method called Basin Hopping [30], which is a hybrid between simulated annealing and gradient descent-based methods.

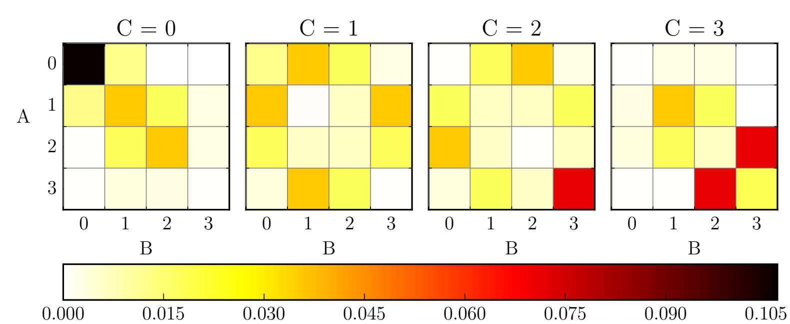


FIG. 8: Symmetrized version of the Fritz distribution ($V_F \simeq 3$).

Despite our varied efforts, it remains unclear whether or not our results share the same possibilistic structure as P_{F} because of the inequality itself or the ill-conditioned nature of the optimization. More likely, this phenomena is due to a combination of both effects.

It is also worth reporting that the distribution in Fig. 6 is not accessible when the bipartite states in Eq. (16) are restricted to maximally entangled states. This result resembles a feature of quantum mechanics originally presented by Methot and Scarani [31] which demonstrates that entanglement and non-classicality are different resources.

The second inequality in consideration is the *symmetric* inequality I_3 (Eq. (15)). This inequality was able to achieve a greater relative violation than I_1 as optimizations converged to a distribution (Fig. 7) with relative violation $V_F \simeq 2.61$. Just as before, the maximum violating distribution shares a possibilistic structure with the Fritz distribution. However, this result is far more surprising. Since the inequality is symmetric with respect to the permutation of parties, one would naively expect that the maximum violating distribution be symmetric as well. If it were not, one could construct a distribution that violates I_3 more by symmetrizing P_{F} by summing over all possible permutations and normalizing. The symmetrized Fritz distribution is plotted in Fig. 8. This distribution is also non-classical with relative violation $V_F = 3$. As it turns out, numerical optimizations suggest that the symmetrized Fritz distribution is *not* quantum-accessible in the sense that is cannot be written as Eq. (16). Therefore we conclude that the space of quantum accessible correlations on the Triangle structure is non-convex; a reiteration of a result mentioned in Wolfe *et al.* [17].

Finally for both I_0 and I_2, the maximum relative violation happens to be $V_F = 1$ (up to numerical precision). This means that the Fritz distribution either represents a local minimum of the parameter space or is in fact *the* maximally violating distribution. Without further investigations, it remains unclear if this behavior is related to the universality of the Fritz distribution or if its an artifact of the methods used to derive I_0 and I_2.

C. Rosset Variation

Todo (TC Fraser): This might take a page

VIII. CONCLUSIONS

In Section IV, we elucidated that compatibility in the Triangle structure has been an unsolved problem for nearly a decade. Initially, Branciard *et al.* [15] identified the Triangle structure as a challenging causal structure for studying non-classicality. Afterwards, Henson *et al.* [16] were able to to classify the Triangle structure as *interesting* which partially explains why it remains an enigma. Recently, Fritz [11] discovered the Fritz distribution as the first example of quantum non-classicality in the Triangle structure. Heretofore, researchers have been able to derive causal incompatibility inequalities [13, 16, 17] but unable to show that any are non-trivial in the context of quantum non-classicality. In Section V the Fritz distribution was described in detail along with Fritz's Problem [11, Problem 2.17].

In Section VI, we presented the first examples of causal compatibility inequalities capable of having quantum violations in the sense that they are violated by quantum accessible distributions. This result was made possible through extensive use of the Inflation Technique (Appendix A) when applied to the Fritz distribution. In the context of Fritz's Problem, our result corresponds to a resolution to the following refinement of Fritz's original Problem.

R.1 Find proofs for causal incompatibility of quantum distributions on the Triangle structure that are *not* reliant on perfect correlations.

Ultimately, this work accomplishes two important tasks. First and foremost, the derivation of inequalities such as Eq. (12) demonstrates the utility of the Inflation technique specifically for quantum causal inference problems. The Inflation Technique was able to derive causal compatibility inequalities that admit quantum violations in the Triangle structure – a task that no other existing techniques have succeeded at. In this sense, the results of Section VI showcase how the Inflation technique is capable of generating new proofs of causal incompatibility specifically of quantum distributions for the Triangle structure. Consequently, this work strongly indicates that the Inflation technique has the potential to solve a assortment of unsolved causal inference problems including those manifest in quantum foundations and quantum information theory. Secondly, Section VII A demonstrates that the derived causal compatibility inequalities are robust to varying degrees of noise, directly revealing a critical departure from Fritz's original proof of the incompatibility of the Fritz's distribution (Section V) which relies on perfect (non-noisy) correlations. In conclusion **R.1** has been positively solved through the derivation of inequalities such as Eq. (12) and those found in **??**.

In consideration of out resolution to **R.1**, we additionally propose two more refinements to Fritz's Problem that are concerned more with the nature of the entanglement resource being exploited in a incompatible quantum distribution:

R.2 Find incompatible quantum distributions on the Triangle structure that require entanglement in more than one shared state.

R.3 Find incompatible quantum distributions on the Triangle structure that satisfy all Bell structure inequalities under Fritz-type embeddings.

Fritz's Problem is concerned with identifying incompatible quantum distributions in the Triangle structure that are qualitatively different than the non-classical distributions of the Bell structure discussed in Section III. When specifically concerned with entanglement resources, Bell structure non-classicality exploits the entanglement of a single bipartite quantum state. In the Triangle structure, the observable nodes A, B, C have access to potentially three different bipartite quantum states. As noted in Section V, the Fritz distribution can be implemented using only a single entangled state[6]. The question, as proposed by **R.2**, remains: does there exist incompatible quantum distributions that genuinely require the entanglement resources of three states? Finding such distributions would affirmatively demonstrate a new form of entanglement resource distinct to the Triangle structure. Moreover **R.3** is justified as a valid interpretation of Fritz's Problem in the following sense: the Fritz distribution was constructed by embedding a quantum distribution incompatible with the Bell structure (Fig. 1) into the Triangle structure via the re-imagined version of the Triangle scenario (Fig. 3). Furthermore, the logic of Section V shows that *any* Bell structure compatibility inequality can be upgraded to a compatibility inequality for the Triangle structure, provided one restricts the domain of those inequalities to distributions possessing appropriate perfect correlations. Suppose one relaxed that restriction, then it becomes qualitatively reasonable to associate incompatibility via violations of upgraded inequalities with the same variety of incompatibility that is familiar in the Bell structure. Therefore, satisfying all such upgraded inequalities is a sufficient condition for a new type of incompatibility that is genuinely distinct from Bell structure incompatibility. In summary, it remains unclear whether or not **R.2** or **R.3** is a harder problem or if either is even feasible.

ACKNOWLEDGMENTS

This research was supported in part by Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is supported by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Research, Innovation and Science.

Appendix A: Summary of Inflation Technique

1. Satisfaction of d-Separation Relations

Todo (TC Fraser): Absorb the relevant components of the next section into this one Todo (TC Fraser): Given a distribution and a causal structure, does the distribution satisfy the conditional d-Separation relations?

2. Marginal Satisfiability and Inequalities

Given a set of observable variables $\mathcal{J} = \{v_1, \ldots, v_k\}$ where each $v \in \mathcal{J}$ is a statistical random variable, one can conceivably make a measurement over of these variables and obtain a joint probability distribution $P_{\mathcal{J}} = P_{v_1,\ldots,v_n}$. More generally however, and depending on the context, it is possible to make a measurement on a set of the system $V \subset \mathcal{J}$ in order to obtain a marginal distribution P_V. With complete generality however, one could in principle be limited so as to only be able to record observations on portions of the system $\mathcal{M} = \{V_1, \ldots, V_m \mid \forall i : V_i \subseteq \mathcal{J}\}$ obtaining a collection of marginal probability distributions $\{P_{V_1}, \ldots, P_{V_m} \mid \forall i : V_i \in \mathcal{M}\}$.

The set $\mathcal{M} = \{V \mid V \subseteq \mathcal{J}\}$ of subsets of $\mathcal{J}$ is referred to as the **marginal scenario** and each element $V \in \mathcal{M}$ is termed a **(marginal) context** of $\mathcal{M}$. The complete set of marginal distributions is referred to as the **marginal model** and is denoted with an superscript $P^{\mathcal{M}}$:

$$P^{\mathcal{M}} = \{P_V \mid V \in \mathcal{M}\}$$

[6] Of course there exists implementations that make use of three entangled states but at minimum only one is necessary.

A marginal model acts as the most general description of a family of observations that can be made over $\mathcal{J}$. Strictly speaking, as defined by [32], a marginal scenario forms an *abstract simplicial complex* where it is required that all subsets of contexts are also contexts.

$$\forall V \in \mathcal{M}, V' \subset V : V' \in \mathcal{M}$$

Throughout this document, we exclusively consider (without loss of generality) the *maximal* marginal scenario; restricting our focus to the largest contexts in the marginal scenario. Additionally, all marginal scenarios are taken to be complete in the sense that the marginal scenario covers the complete set of observable variables $\mathcal{J}$.

$$\mathcal{J} = \bigcup_{V \in \mathcal{M}} V$$

It is important to note that a set of causal parameters is not unique; the numerical nature of set of causal parameters is a choice. Altogether, a marginal model $P^{\mathcal{M}}$ is **compatible** with a causal structure $\mathcal{G}$ if there exists a choice of causal parameters such that each context $P_V \in P^{\mathcal{M}}$ can be *recovered* from the following series of operations:

1. Obtain a joint probability distribution over *all* variables of the causal structure using the chosen casual parameters:[7]

2. Marginalize over the latent variables of $\mathcal{G}$ to obtain a joint distribution over the observed variables $\mathcal{N}_O$:

$$P_{\mathcal{N}_O} = \sum_{\mathcal{N}_L} P_{\mathcal{N}} \tag{A1}$$

3. Finally marginalize over the observed nodes not in V to obtain P_V:

$$P_V = \sum_{\mathcal{N}_O \setminus V} P_{\mathcal{N}_O} \tag{A2}$$

It is important to note that compatibility between $P^{\mathcal{M}}$ and $\mathcal{G}$ requires that it is possible to recover *each* context $P_V \in P^{\mathcal{M}}$ from the *same* set of causal parameters. If it is not possible for $P^{\mathcal{M}}$ to be recovered for *any* choice of causal parameters, then $P^{\mathcal{M}}$ is **incompatible**. In Section II, we defined the notion of a causal compatibility inequality without providing a means for deriving/computing them. This section aims to delineate algorithms and methods that can be used to derive causal compatibility inequalities for generic causal inference problems. Recalling the definition of causal compatibility presented in Section II, a marginal model $P^{\mathcal{M}}$ is compatible with a causal structure $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ if it can be recovered from a series of operations. First choose a set of causal parameters $\left\{P_{n|\mathsf{Pa}_{\mathcal{G}}(n)} \mid n \in \mathcal{N}\right\}$ and then construct a joint distribution $P_{\mathcal{N}} = \prod_{n \in \mathcal{N}} P_{n|\mathsf{Pa}_{\mathcal{G}}(n)}$ over all nodes $\mathcal{N}$ (this step is Eq. (2)). Afterwards, each marginal distribution $P_V \in \mathcal{M}$ should be recoverable from a marginalization over $P_{\mathcal{N}}$ (this step is the combination of Eqs. (A1,A2)). Intuitively, these two requirements for a compatible marginal model $P^{\mathcal{M}}$ provide two necessary components of compatibility:

1. The marginal model $P^{\mathcal{M}}$ must admit a joint distribution $P_{\mathcal{N}}$ (Eqs. (A1,A2)).

2. The joint distribution $P_{\mathcal{N}}$ must satisfy the Markov separability criteria (Eq. (2)).

In practice, this is the natural order in which to classify the compatibility of a marginal model $P^{\mathcal{M}}$. First check if the marginal model admits a joint distribution and if and only if it does, determine whether or not the local Markov property can be satisfied for some choice of causal parameters. Intuitively, there are two reasons why a given marginal model $P^{\mathcal{M}}$ would be deemed incompatible. One possibility is that the marginal model $P^{\mathcal{M}}$ does not admit a joint distribution $P_{\mathcal{N}}$. Alternatively, it could be possible that a joint distribution $P_{\mathcal{N}}$ exists but can not factor into causal parameters as per Eq. (2).

Colloquially, the first necessary condition, namely the existence (or non-existence) of a joint distribution, is referred to as **the marginal problem**. The marginal problem asks: given a marginal model $P^{\mathcal{M}} = \{P_V \mid V \in \mathcal{M}\}$ marginal

---

[7] Here $\prod$ denotes the product distribution $\prod_{n \in \mathcal{N}} P_{n|\mathsf{Pa}_{\mathcal{G}}(n)} =$.

to the joint variables $\mathcal{J} = \bigcup_{V\in\mathcal{M}} V$, does there exist a joint distribution $P_{\mathcal{J}}$ such that each context P_V can be obtained by marginalizing $P_{\mathcal{J}}$?

$$\forall V \in \mathcal{M} : P_V = \sum_{\mathcal{J}\setminus V} P_{\mathcal{J}} \tag{A3}$$

Note that we have changed notation from the nodes of a causal structure $\mathcal{N}$ to a joint set of variables $\mathcal{J}$. This is done simply because the marginal problem is well-defined independent of a given causal structure. In fact, if a marginal model $P^{\mathcal{M}}$ does not admit a joint distribution, then it cannot be compatible with *any* causal structure whatsoever. A marginal model is said to be *contextual* if it *does not* admit a joint distribution and *non-contextual* otherwise. The marginal problem has applications to numerous areas of mathematics including game theory [33], database theory, knowledge integration of expert system, and of course, quantum information theory [32].

The marginal problem is distinct from the second necessary condition, which we term the **Markov separability problem** in numerous ways. First of all, the marginal problem as applied to causal inference tasks is only concerned with the observable variables $\mathcal{N}_O$ of $\mathcal{G}$. This differs from the Markov separability problem which depends directly on the entire graphical structure of $\mathcal{G}$ and thus must include the latent variables $\mathcal{N}_L$. The purpose of latent variables is to have unspecified behavior which makes checking the local Markov property exceedingly difficult in comparison to the marginal problem. Secondly, Eq. (A3) is inherently a linear system of constraints which be solved rather efficiently using linear programs (to be discussed in Appendix A 2). The local Markov property concerns itself with probabilistic independence constraints which are polynomial. Fortunately, the Inflation Technique (summarized in Appendix A) is able to cast (at least partially) the Markov separability problem as a marginal problem over an inflated causal structure. As such the Inflation Technique in many ways can be considered a relaxation technique which turns polynomial constraints into linear constraints.

In consideration of these observations, we will now endeavor to discuss how causal compatibility inequalities can be derived by casting the marginal problem as a linear program. After doing, it becomes possible to discuss existing methods for deriving constraints on the set of contextual marginal models.

Heretofore, our discussions of compatibility and the marginal problem have been general in the sense of making no reference to the nature the random variables $v \in \mathcal{J}$. In particular, whether the outcomes of v are discrete or continuous, finite or unbounded, or otherwise. In the case of infinite outcomes, the marginal problem can be used to derive entropic inequalities defined over the mutual information between variables $v \in \mathcal{J}$ [32]. Entropic inequalities however have been shown to be less constraining than inequalities defined over variables with finite cardinality [17]. Throughout this document, we restrict our attention to finite outcomes. To assist with the subsequent analysis, we recall some familiar notions of probability theory.

To every discrete random variable v there corresponds a prescribed set of **outcomes** O_v. We also define the set of all **events over** v $\mathcal{S}(v)$[8] to be the set of all functions $s : \{v\} \to O_v$ each representing the event that a measurement on v was made where $s(v) \in O_v$ was observed.

$$\mathcal{S}(v) \equiv \{s : \{v\} \to O_v\}$$

Evidently, $\mathcal{S}(v)$ and O_v have a one-to-one correspondence and this distinction can be confounding. There is rarely any harm in referring synonymously to either as outcomes. Nonetheless, a sheaf-theoretic treatment of contextuality [34] demands the distinction. Specifically for this work, the distinction becomes essential for our discussion and exploit of marginal symmetries in Appendix A 5. As a natural generalization we define the event over a collection of random variables $V = \{v_1, \ldots, v_n\}$ in a parallel manner,

$$\mathcal{S}(V) \equiv \{s : V \to O_V \mid \forall i : s(v_i) \in O_{v_i}\} \tag{A4}$$

Each event s can be compactly represented as a set of mappings over each element of V.

$$s = \{v_1 \mapsto s(v_1), \ldots, v_k \mapsto s(v_k)\} = \{v_i \mapsto s(v_i)\}_{i=1}^{k}$$

The domain $\mathcal{D}(s)$ of an event s is the set of random variables it valuates, i.e. if $s \in \mathcal{S}(V)$, then $\mathcal{D}(s) = V$. Under this framework, a probability distribution P_V can be considered as a map from $\mathcal{S}(V)$ to the unit interval $[0, 1]$.

Example 1. Let $s \in \mathcal{S}(V)$ be an event over V, a probability distribution P_V over V will assign a non-negative probability $P_V(s)$ to s. Let $V = \{v_1, v_2\}$ with each $v \in V$ having 3 distinct outcomes $O_v = \{0, 1, 2\}$. Furthermore let $f \in \mathcal{S}(V)$ be the event $(v_1 \mapsto 2, v_2 \mapsto 1)$. The following notations are interchangeable,

$$P(s) = P_{\mathcal{D}(s)}(s) = P_V(s) = P_{v_1,v_2}(s) = P(v_1 \mapsto s_1(v_1), v_2 \mapsto s_2(v_2)) = P(v_1 = 2, v_2 = 1) = P_{v_1 v_2}(21)$$

Each of which should be read as "the probability that v_1 got outcome 2 and v_2 got outcome 1"

[8] In the language of sheaf-theory, $\mathcal{S}(v)$ is the *sheaf of events* [34].

The marginal problem inherently depends on the concept of probabilistic marginalization. This concept can be understood at the level of events; one event $s \in \mathcal{S}(V)$ can be "marginalized" or restricted to a smaller event $s' \in \mathcal{S}(W)$ whenever $W \subseteq V$.

Definition 2. For every $W \subseteq V$ and $s \in \mathcal{S}(V)$, the **restriction of s onto W** (denoted $s|_W \in \mathcal{S}(W)$) is the event in $\mathcal{S}(W)$ that *agrees* with each of s's assignments for variables in W.

$$\forall v \in W : s|_W(v) = s(v)$$

For every marginal scenario $\mathcal{M} = \{V_1, \ldots, V_k\}$, it is useful to put special emphasis on the **joint events** $\mathcal{S}(\mathcal{J})$ which represent all possible global events over the entire set of joint variables. Similarly, we define the **context events** for a particular context $V \in \mathcal{M}$ as $\mathcal{S}(V)$. Finally, we elect to define the **marginal events** as the disjoint union of all context events and by an abuse of notation we will denote this disjoint union as $\mathcal{S}(\mathcal{M})$.

$$\mathcal{S}(\mathcal{M}) = \coprod_{V \in \mathcal{M}} \mathcal{S}(V)$$

Each marginal section $m \in \mathcal{S}(\mathcal{M})$ has a domain $\mathcal{D}(m) = V$ for some $V \in \mathcal{M}$. By construction each marginal event $m \in \mathcal{S}(\mathcal{M})$ is a restriction of some joint event $j \in \mathcal{S}(\mathcal{J})$. The marginalization operation (Eq. (A3)) of the marginal problem is a linear operator. This linear operator maps a joint probability distribution $P_{\mathcal{J}} : \mathcal{S}(\mathcal{J}) \to [0,1]$ into a marginal model $P^{\mathcal{M}} = \{P_V : \mathcal{S}(V) \to [0,1] \mid V \in \mathcal{M}\}$. Since $\mathcal{S}(\mathcal{M})$ and $\mathcal{S}(\mathcal{J})$ are finite, the marginal problem can be casted as a $|\mathcal{S}(\mathcal{M})| \times |\mathcal{S}(\mathcal{J})|$ matrix.

Definition 3. The **incidence matrix** M for a marginal scenario $\mathcal{M} = \{V_1, \ldots, V_k\}$ is a bit-wise matrix where the columns are indexed by *joint* events $j \in \mathcal{S}(\mathcal{J})$ and the rows are events by *marginal* events $m \in \mathcal{S}(\mathcal{M})$. The entries of M are populated whenever a marginal event m is a restriction of the joint event j.

$$M_{m,j} \equiv \begin{cases} 1 & m = j|_{\mathcal{D}(m)} \\ 0 & \text{otherwise} \end{cases}$$

The incidence matrix has $|\mathcal{S}(\mathcal{J})|$ columns, $|\mathcal{S}(\mathcal{M})| = \sum_i |\mathcal{S}(V_i)|$ rows and $k|\mathcal{S}(\mathcal{J})|$ non-zero entries.

To illustrate this concretely, consider the following example:

Example 4. Let $\mathcal{J}$ be 3 binary variables $\mathcal{J} = \{A, B, C\}$ and $\mathcal{M}$ be the marginal scenario $\mathcal{M} = \{\{A,B\},\{B,C\},\{A,C\}\}$. The incidence matrix becomes:

$$M = \begin{array}{r c}
(A,B,C) \mapsto & \begin{array}{cccccccc} (0,0,0) & (0,0,1) & (0,1,0) & (0,1,1) & (1,0,0) & (1,0,1) & (1,1,0) & (1,1,1) \end{array} \\
\begin{array}{r}
(A\mapsto 0, B\mapsto 0) \\ (A\mapsto 0, B\mapsto 1) \\ (A\mapsto 1, B\mapsto 0) \\ (A\mapsto 1, B\mapsto 1) \\ (B\mapsto 0, C\mapsto 0) \\ (B\mapsto 0, C\mapsto 1) \\ (B\mapsto 1, C\mapsto 0) \\ (B\mapsto 1, C\mapsto 1) \\ (A\mapsto 0, C\mapsto 0) \\ (A\mapsto 0, C\mapsto 1) \\ (A\mapsto 1, C\mapsto 0) \\ (A\mapsto 1, C\mapsto 1)
\end{array} &
\begin{pmatrix}
\mathbf{1} & \mathbf{1} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \mathbf{1} & \mathbf{1} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \mathbf{1} & \mathbf{1} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \mathbf{1} & \mathbf{1} \\
\mathbf{1} & 0 & 0 & 0 & \mathbf{1} & 0 & 0 & 0 \\
0 & \mathbf{1} & 0 & 0 & 0 & \mathbf{1} & 0 & 0 \\
0 & 0 & \mathbf{1} & 0 & 0 & 0 & \mathbf{1} & 0 \\
0 & 0 & 0 & \mathbf{1} & 0 & 0 & 0 & \mathbf{1} \\
\mathbf{1} & 0 & \mathbf{1} & 0 & 0 & 0 & 0 & 0 \\
0 & \mathbf{1} & 0 & \mathbf{1} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \mathbf{1} & 0 & \mathbf{1} & 0 \\
0 & 0 & 0 & 0 & 0 & \mathbf{1} & 0 & \mathbf{1}
\end{pmatrix}
\end{array} \tag{A5}$$

The incidence matrix needs to act (to the right) on a vector representing the joint probability $P_{\mathcal{J}}$ distribution and output a vector representing the marginal model $P^{\mathcal{M}}$.

Definition 5. The **joint distribution vector** $\mathcal{P}^{\mathcal{J}}$ for a probability distribution $P_{\mathcal{J}}$ is the vector whose entries are the positive, real-valued probabilities that $P_{\mathcal{J}}$ assigns to each joint event $j \in \mathcal{S}(\mathcal{J})$. $\mathcal{P}^{\mathcal{J}}$ shares the same indices as the *column* indices of M.

$$\forall j \in \mathcal{S}(\mathcal{J}) : \mathcal{P}^{\mathcal{J}}_j = P_{\mathcal{J}}(j)$$

Definition 6. Analogously, **marginal distribution vector** $\mathcal{P}^{\mathcal{M}}$ for a marginal model $P^{\mathcal{M}} = \{P_{V_1}, \ldots, P_{V_k}\}$ is the vector whose entries are probabilities over the set of marginal outcomes $\mathcal{S}(\mathcal{M})$. $\mathcal{P}^{\mathcal{M}}$ shares the same indices as the *row* indices of M.

$$\forall m \in \mathcal{S}(\mathcal{M}) : \mathcal{P}^{\mathcal{M}}_m = P_{\mathcal{D}(m)}(m)$$

By design, the marginal and joint distribution vectors are related via the incidence matrix M. Given a joint distribution vector $\mathcal{P}^{\mathcal{J}}$ one can obtain the marginal distribution vector $\mathcal{P}^{\mathcal{M}}$ by multiplying M by $\mathcal{P}^{\mathcal{J}}$.

$$\mathcal{P}^{\mathcal{M}} = M \cdot \mathcal{P}^{\mathcal{J}} \tag{A6}$$

Because the marginal distribution vector $\mathcal{P}^{\mathcal{M}}$ is simply another representation of the marginal model $P^{\mathcal{M}}$, we will henceforth use the language interchangeably. This is likewise done so for the joint distribution (vector) $P_{\mathcal{J}}$ ($\mathcal{P}^{\mathcal{J}}$). As a quick remark, the particular ordering of the rows and columns of M carries *no importance*, but it *must* be consistent between M, $\mathcal{P}^{\mathcal{J}}$ and $\mathcal{P}^{\mathcal{M}}$. The marginal problem can now be rephrased in the language of the incidence matrix. Suppose one obtains a marginal distribution vector $\mathcal{P}^{\mathcal{M}}$; the marginal problem becomes equivalent to the question: *Does there exist a joint distribution vector $\mathcal{P}^{\mathcal{J}}$ such that Eq. (A6) holds?*

The marginal problem can sometimes take on a more general variant that does not begin with a specific marginal model [32, 35, 36]: *Given a marginal scenario $\mathcal{M}$, what is the set of all non-contextual marginal models?* Pitowsky [37] demonstrates that the set of non-contextual marginal models forms a *convex* polytope called the **marginal polytope**. The extremal rays of a marginal polytope directly correspond to the columns of M which correspond to each of the $|\mathcal{S}(\mathcal{J})|$ *deterministic* joint distributions $\mathcal{P}^{\mathcal{J}}$. Since all joint distributions $\mathcal{P}^{\mathcal{J}}$ are probability distributions, their entries must sum to unity.

$$\sum_{j \in \mathcal{S}(\mathcal{J})} \mathcal{P}^{\mathcal{J}}_j = 1$$

This normalization defines the convexity of the polytope; all non-contextual marginal models are convex mixtures of the deterministic marginal models pursuant to Eq. (A6). The marginal polytope is a beneficial tool for understanding contextuality. First, the *facets* of a marginal polytope correspond to a finite set of linear inequalities that are complete in the sense that all contextual distributions violate at least one facet inequality [10]. From the perspective of a marginal polytope, convex hull algorithms can be used to compute a representation of the complete set of linear inequalities and completely solve the marginal problem. Alternatively, one can completely characterize the separation between the space of contextual and non-contextual distributions by the system of inequalities associated with Eq. (A6). Explicitly,

$$\begin{aligned} \forall m \in \mathcal{S}(\mathcal{M}) : &\quad \mathcal{P}^{\mathcal{M}}_m - \sum\nolimits_j M_{m,j} \mathcal{P}^{\mathcal{J}}_j \geq 0 \\ \forall m \in \mathcal{S}(\mathcal{M}) : &\quad -\mathcal{P}^{\mathcal{M}}_m + \sum\nolimits_j M_{m,j} \mathcal{P}^{\mathcal{J}}_j \geq 0 \\ \forall j \in \mathcal{S}(\mathcal{J}) : &\quad \mathcal{P}^{\mathcal{J}}_j \geq 0 \end{aligned} \tag{A7}$$

Geometrically Eq. (A7) is a family of $2|\mathcal{S}(\mathcal{M})| + |\mathcal{S}(\mathcal{J})|$ half-planes bounding $\mathbb{R}^{|\mathcal{S}(\mathcal{M})|} \times \mathbb{R}^{|\mathcal{S}(\mathcal{J})|}$. However testing the contextuality of a marginal model $\mathcal{P}^{\mathcal{M}}$ requires a collection of inequalities bounding just $\mathbb{R}^{|\mathcal{S}(\mathcal{M})|}$. Transforming Eq. (A7) into the desired system of inequalities is the task of linear-quantifier elimination. A popular tool for linear quantifier elimination is **Fourier-Motzkin elimination** [17, 35, 38, 39]. Applying the Fourier-Motzkin procedure to completely solve the marginal problem is discussed in more detail in Fritz and Chaves [32]. An excellent survey of existing techniques for solving the marginal problem including Equality Set Projection [39] and Hardy-type hypergraph transversals can be found in Wolfe *et al.* [17]. In conclusion, there are a number of computational tools available to solve the marginal problem completely whenever no marginal model is provided.

Each of the above mentioned techniques suffers from computational complexity limitations. For example, the Fourier-Motzkin procedure is in the worst case doubly exponential in the number of initial inequalities [38]. For the purposes of this research, solving the marginal problem without reference to a marginal model was intractable. This will become apparent in Section VI when the inflation technique is applied to the Triangle structure, producing marginal models that are very large. Moreover, algorithms to partially solve the marginal problem such as Hardy-type hypergraph transversals were used to determine hundreds of thousands of causal compatibility inequalities for the Triangle structure; none of which were able to be violated quantumly using numerical optimizations. Therefore, simpler approaches to the marginal problem needed to be considered.

Luckily, the Fritz distribution (Section V) allows one to avoid the complexity issues of the complete marginal problem and instead focus on the original problem of determining where or not a particular marginal model admitted a joint distribution or not. This question is naturally framed as the following linear problem:

Definition 7. **The Marginal Linear Program** is the following linear program:

$$\begin{aligned} \text{minimize:} \quad & \emptyset \cdot x \\ \text{subject to:} \quad & x \succeq 0 \\ & M \cdot x = \mathcal{P}^{\mathcal{M}} \end{aligned}$$

If this "optimization"[9] is *feasible*, then there exists a vector x than can satisfy Eq. (A6) and is a valid joint distribution vector. Therefore feasibility implies that $\mathcal{P}^{\mathcal{J}}$ exists and is precisely x. Moreover if the marginal linear program is *infeasible*, then there *does not* exist a joint distribution $\mathcal{P}^{\mathcal{J}}$.

To every linear program, there exists a dual linear program that characterizes the feasibility of the original [40]. Constructing the dual linear program is straightforward [41].

Definition 8. **The Dual Marginal Linear Program**:

$$\begin{aligned} \text{minimize:} \quad & y \cdot \mathcal{P}^{\mathcal{M}} \\ \text{subject to:} \quad & y \cdot M \succeq 0 \end{aligned}$$

Where y is a real valued vector with the same length as $\mathcal{P}^{\mathcal{M}}$.

The dual marginal linear program not only answers the marginal problem for a specific marginal model $\mathcal{P}^{\mathcal{M}}$ but as a by-product provides an inequality that witnesses its contextuality. If this is not obvious, first notice that the dual problem is *never infeasible*; by choosing y to be trivially the null vector $\emptyset$ of appropriate size, all constraints become satisfied. Secondly if the dual constraint $y \cdot M \succeq 0$ holds *and* the primal is feasible, then the following must hold:

$$y \cdot \mathcal{P}^{\mathcal{M}} = y \cdot M \cdot x \geq 0 \tag{A8}$$

Therefore the *sign* of the dual objective $d \equiv \min\bigl(y \cdot \mathcal{P}^{\mathcal{M}}\bigr)$ classifies a marginal model's contextuality. If $d < 0$ then Eq. (A8) is violated and therefore $\mathcal{P}^{\mathcal{M}}$ is contextual. Likewise if $d \geq 0$ (satisfying Eq. (A8)), then $\mathcal{P}^{\mathcal{M}}$ is non-contextual.[10] This is manifestation of Farkas's lemma [40].

An **infeasibility certificate** [43] is any vector y that satisfies the constraint of Theorem 8 and also permits violation of Eq. (A8) for some marginal distribution vector $\mathcal{P}^{\mathcal{M}}$.

$$y \in \mathbb{R}^{|\mathcal{S}(\mathcal{M})|} : y \cdot M \succeq 0, \quad y \cdot \mathcal{P}^{\mathcal{M}} < 0$$

Most linear program software packages such as Mosek [44], Gurobi [45], CPLEX [46] and CVX/CVX OPT [47, 48] are capable of producing infeasibility certificates. Furthermore, for every y satisfying $y \cdot M \succeq 0$ there corresponds a **certificate inequality** that constraints the set of non-contextual marginal models. If y is an infeasibility certificate, then all non-contextual marginal models $\mathcal{P}^{\mathcal{M}}$ must satisfy Eq. (A8) making $y \cdot \mathcal{P}^{\mathcal{M}} \geq 0$ a valid causal compatibility inequality.

3. Inflation Technique

The **Inflation Technique**, invented by Wolfe *et al.* [17] and inspired by the *do calculus* and *twin networks* of Pearl [8], is a family of causal inference techniques that can be used to determine if a marginal model $P^{\mathcal{M}}$ is compatible or incompatible with a given causal structure $\mathcal{G}$. As a preliminary summary, the inflation technique begins by *augmenting* a causal structure with additional copies of its nodes, producing an *inflated* causal structure, and then exposes how causal inference tasks on the inflated causal structure can be used to make inferences on the original causal structure. As a example, inflations of the Triangle structure are depicted in Fig. 9. Copies of nodes in the inflated causal structure are distinguished by an additional subscript called the **copy-index**. For example node A of Fig. 2 has copies A_1, A_2, A_3, A_4 in the *Web inflation* of Fig. 9d. All such copies are deemed equivalent via a **copy-index**

---

[9] "Optimization" is presented in quotes here because the minimization objective is trivially always zero ($\emptyset$ denotes the null vector of all zero entries). The primal value of the linear program is of no interest, all that matters is its *feasibility*.

[10] Actually, if $d \geq 0$ then it is exactly $d = 0$ due to the existence of the trivial $y = \emptyset$. This observation is an instance of the *Complementary Slackness Property* of [42]. Moreover, if $d < 0$, then it is unbounded $d = -\infty$. This latter point becomes clear upon recognizing that for any y with $d < 0$, another $y' = \alpha y$ can be constructed (with $\alpha > 1$) such that $d' = \alpha d < d$. Since a more negative d' can always be found, it must be that d is unbounded. This is a demonstration of the fundamental *Unboundedness Property* of [42]; if the dual is unbounded, then the primal is infeasible.

equivalence relation denoted '$\sim$'. A copy-index is effectively arbitrary, so we will refer to an arbitrary inflated copy of A as A'.

$$A \sim A_1 \sim A' \not\sim B \sim B_1 \sim B'$$

We preemptively generalize the notion of copy-index equivalence to other mathematical objects like sets, graphs, and groups by saying that $X \sim Y$ if and only if $X = Y$ upon removal of the copy-index. Equipped with the common graph-theoretic terminology and notation presented in Section II, an inflation can be formally defined as follows:

Todo (TC Fraser): Incorporate these definitions into the discussion

Definition 9. An **induced subgraph** of $\mathcal{G}$ for a subset of nodes $N \subseteq \mathcal{N}$ is another graph composed of nodes N and all edges e of the original graph that are contained in N. An **ancestral subgraph** of $\mathcal{G}$ for a subset of nodes $N \subseteq \mathcal{N}$ is the induced subgraph due to the ancestry of N.

$$\mathsf{AnSub}_{\mathcal{G}}(N) \equiv \mathsf{Sub}_{\mathcal{G}}\big(\mathsf{An}_{\mathcal{G}}(N)\big)$$

Definition 10. An **inflation** of a causal structure $\mathcal{G}$ is another causal structure $\mathcal{G}'$ such that:

$$\forall n' \in \mathcal{N}' : \mathsf{AnSub}_{\mathcal{G}'}(n') \sim \mathsf{AnSub}_{\mathcal{G}}(n) \tag{A9}$$

Verbally, for each inflated node n', the ancestral sub-graph of n' in $\mathcal{G}'$ is equivalent to its counterpart in $\mathcal{G}$ up to copy-index.

The motivation behind this definition is that if the ancestry of each node n' in $\mathcal{G}'$ plays the *exact same* role as the ancestry of its source copy n in $\mathcal{G}$, then every compatible joint distribution $P_{\mathcal{N}}$ on $\mathcal{G}$ can be used to *create* compatible joint distributions on $\mathcal{G}'$.

To do this, first notice that every compatible joint distribution $P_{\mathcal{N}}$ uniquely defines a set of causal parameters for $\mathcal{G}$. It is possible to obtain these causal parameters by computing the single-variable marginal distributions of $P_{\mathcal{N}}$,

$$\forall n \in \mathcal{N} : P_{n|\mathsf{Pa}_{\mathcal{G}}(n)} = \sum_{\mathcal{N}\setminus n} P_{\mathcal{N}} \tag{A10}$$

Now it is possible to make use of the inflation criteria (Eq. (A9)) which at least enforces that $\mathsf{Pa}_{\mathcal{G}'}(n') \sim \mathsf{Pa}_{\mathcal{G}}(n)$. From this, one can create a set of **inflated causal parameters** using Eq. (A10),

$$\forall n' \in \mathcal{N}' : P_{n'|\mathsf{Pa}_{\mathcal{G}'}(n')} \equiv P_{n|\mathsf{Pa}_{\mathcal{G}}(n)} \tag{A11}$$

Which in turn uniquely defines a compatible joint distribution $P_{\mathcal{N}'}$ on $\mathcal{G}'$[11].

$$P_{\mathcal{N}'} = \prod_{n' \in \mathcal{N}'} P_{n'|\mathsf{Pa}_{\mathcal{G}'}(n')} \tag{A12}$$

Together, Eqs. (A10-A12) form the **weak inflation lemma**: compatible joint distributions over $\mathcal{N}$ induce compatible joint distributions over $\mathcal{N}'$. Before generalizing to the inflation lemma, a careful observation needs to be made. For any pair of subsets of nodes $N \subseteq \mathcal{N}$ and $N' \subseteq \mathcal{N}'$ that are equivalent up to copy-index $N \sim N'$, if $\mathsf{AnSub}_{\mathcal{G}}(N) \sim \mathsf{AnSub}_{\mathcal{G}'}(N')$ then any compatible *marginal* distribution P_N, where $N \subset \mathcal{N}$, induces a compatible marginal distribution $P_{N'}$ over N'. In fact, N' must contain no duplicate nodes up to copy index and thus $P_N = P_{N'}$. These subsets of nodes are so essential to the Inflation Technique they are assigned a special name: sets $N' \subseteq \mathcal{N}'$ as called the **injectable sets** of $\mathcal{G}'$ and $N \subset \mathcal{N}$ the **images of the injectable sets** of $\mathcal{G}$.

$$\begin{aligned}\mathsf{Inj}_{\mathcal{G}}(\mathcal{G}') &\equiv \{N' \subseteq \mathcal{N}' \mid \exists N \subseteq \mathcal{N} : N \sim N'\} \\ \mathsf{ImInj}_{\mathcal{G}}(\mathcal{G}') &\equiv \{N \subseteq \mathcal{N} \mid \exists N' \subseteq \mathcal{N}' : N \sim N'\}\end{aligned}$$

Injectable sets are incredibly useful for causal compatibility because they allow one to extend the weak inflation lemma to *the* inflation lemma:

[11] Of course, there exists compatible joint distributions on $\mathcal{N}'$ that can not be constructed in this manner. The claim is that all joint distributions that are constructed in this way are compatible.

Lemma 11. ***The Inflation Lemma*** *[17, Lemma 3] Given a particular inflation $\mathcal{G}'$ of $\mathcal{G}$ and a marginal model $\{P_N \mid N \in \mathsf{ImInj}_{\mathcal{G}}(\mathcal{G}')\}$ compatible with $\mathcal{G}$ then all marginal models $\{P_{N'} \mid N' \in \mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')\}$ are compatible with $\mathcal{G}'$ provided that $P_N = P_{N'}$ for all instances where $N \sim N'$.*

Diagrammatically, Lemma 11 can be visualized as follows:

$$\begin{tikzcd}
\underbrace{\{P_N \mid N \in \mathsf{ImInj}_{\mathcal{G}}(\mathcal{G}')\}}_{\text{compatible with } \mathcal{G}} \arrow[r] & \left\{P_{n|\mathsf{Pa}_{\mathcal{G}}(n)} \mid n \in \mathcal{N}\right\} \arrow[d, "\text{define}"] \\
\underbrace{\{P_{N'} \mid N' \in \mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')\}}_{\text{compatible with } \mathcal{G}'} & \left\{P_{n'|\mathsf{Pa}_{\mathcal{G}'}(n')} \mid n' \in \mathcal{N}'\right\} \arrow[l]
\end{tikzcd}$$

The inflation lemma is the most important result of the causal inflation technique. The contrapositive version of Lemma 11 is a powerful tool for deriving causal compatibility constraints. Any causal compatibility constraint, such as a causal compatibility inequality, that constrains injectable marginal models $\{P_{N'} \mid N' \in \mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')\}$ can be *deflated* into a valid compatibility constraint on marginal models $\{P_N \mid N \in \mathsf{ImInj}_{\mathcal{G}}(\mathcal{G}')\}$ of the original causal structure by dropping all references to the copy-indices. Additionally, the inflation lemma holds when considering any subset of $\mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')$ (analogously $\mathsf{ImInj}_{\mathcal{G}}(\mathcal{G}')$). Therefore, one only needs to consider injectable sets that are composed of *observable nodes*.

Heretofore, the Inflation Technique as stated by Lemma 11 does not facilitate the derivation of any non-trivial inequalities. The Inflation lemma simply casts causal inference problems on $\mathcal{G}$ to causal inference problems on $\mathcal{G}'$. Fortunately, an inflated causal structure $\mathcal{G}'$ possesses its own conditional independence relations which can be exploited to turn linear inequalities on $\mathcal{G}'$ into polynomial inequalities on $\mathcal{G}$.

Suppose that one were to obtain an inequality on $\mathcal{G}'$ that contained probabilistic terms such as $P_{N'}$ where $N' \notin \mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')$. At this point, the Inflation lemma is not applicable because it requires probabilistic constraints to consist entirely of distributions $P_{N'}$ where N' is injectable. If for example, $N' \notin \mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')$ could be partitioned into two injectable subsets $N'_1 \in \mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')$ and $N'_2 \in \mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')$ that where **ancestrally independent** in $\mathcal{G}'$,

$$N'_1 \perp N'_2 \iff \mathsf{An}_{\mathcal{G}'}(N'_1) \cap \mathsf{An}_{\mathcal{G}'}(N'_2) = \emptyset$$

Then the compatible marginal context $P_{N'}$ must be probabilistically separable $P_{N'} = P_{N'_1} P_{N'_2}$ due to Eq. (1). Since N'_1 and N'_2 are both injectable, the inflation lemma now applies and the original inequality can be deflated.

$$\begin{tikzcd}
P_{N'} \arrow[r, "N'_1 \cup N'_2 = N", "N'_1 \perp N'_2"'] \arrow[d, dash] & P_{N'_1} P_{N'_2} \arrow[d, dash] \\
N' \notin \mathsf{Inj}_{\mathcal{G}}(\mathcal{G}') \arrow[d, "\times"] & N'_1, N'_2 \in \mathsf{Inj}_{\mathcal{G}}(\mathcal{G}') \arrow[d] \\
?? & P_{N_1} P_{N_2}
\end{tikzcd}$$

Not all sets $N' \subseteq \mathcal{N}'$ can be written as the disjoint union of injectable sets, but those that can are called the **pre-injectable sets**. A set $N' \subseteq \mathcal{N}'$ is a pre-injectable set if it can be decomposed into the disjoint union of injectable sets $N' = \coprod_i N'_i \mid N'_i \in \mathsf{Inj}_{\mathcal{G}}(\mathcal{G}')$ and all pairs N'_i, N'_j are ancestrally independent: $\forall i, j : N'_i \perp N'_j$.

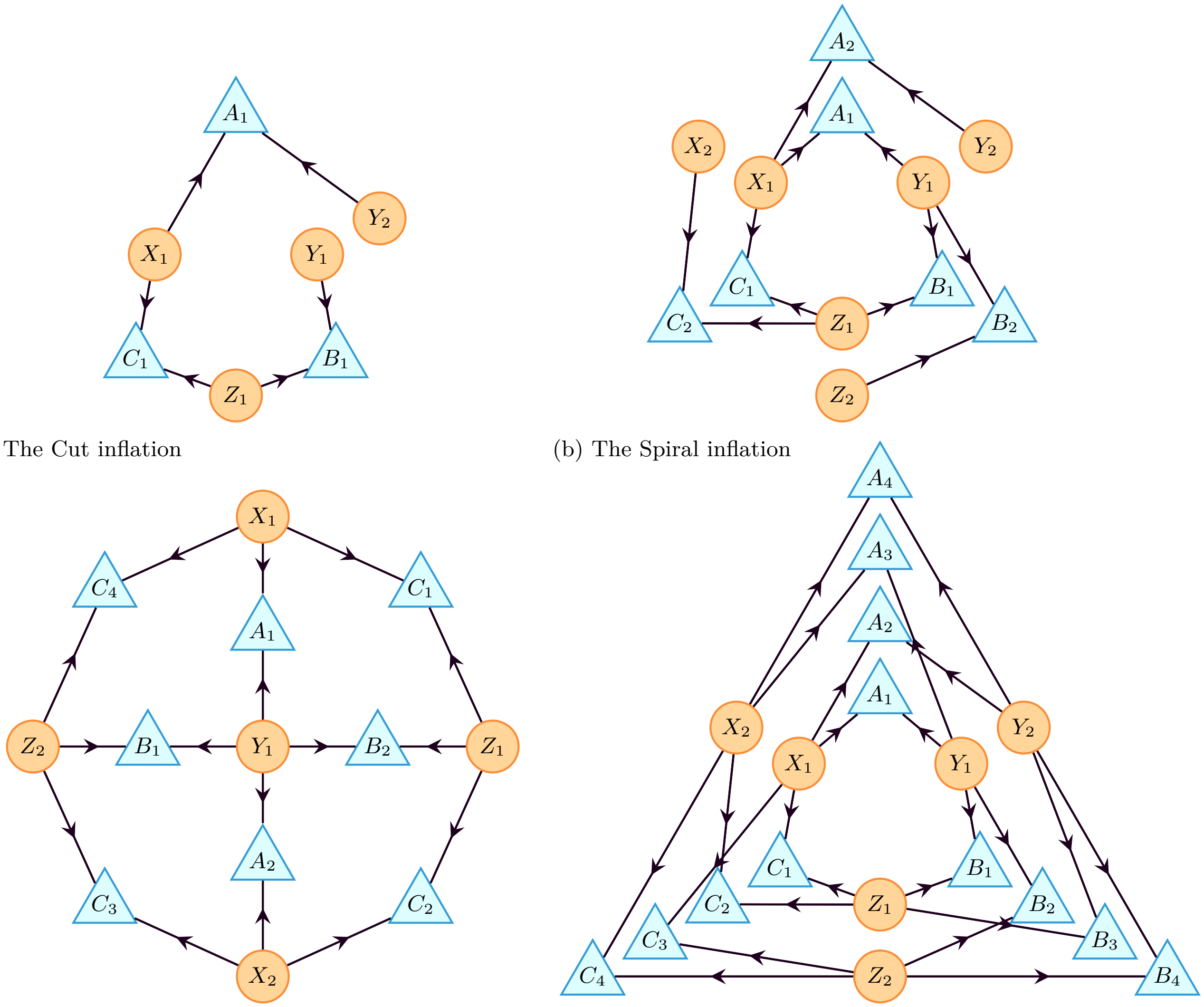
A pre-injectable set N' is of primal importance to the inflation technique because any distribution over N' will factorize according to graphical d-separation conditions [8],

$$P_{N'} = P_{\coprod_i N'_i} = \prod_i P_{N'_i}$$

Application of the inflation lemma to linear inequalities constraining pre-injectable distributions $P_{N'}$ transforms those inequalities into polynomial inequalities over the injectable distributions, allowing one to replace all terms with probability distributions defined over variables in $\mathcal{N}$. Throughout the body of this work, we let $\mathsf{PreInj}_{\mathcal{G}}(\mathcal{G}')$ denote the set of all pre-injectable sets. Naturally, the pre-injectable sets will define a marginal scenario $\mathcal{M}$ that are used to derive compatibility inequalities in Appendix A 3[12].

---

[12] Analogously for a marginal scenario $\mathcal{M}$, the pre-injectable sets $\mathsf{PreInj}_{\mathcal{G}}(\mathcal{G}')$ form an *abstract simplicial complex*. Therefore, in practice, it is completely sufficient to focus on the maximal pre-injectable sets



(a) The Cut inflation

(b) The Spiral inflation

(c) The Wagon-Wheel inflation

(d) The Web Inflation

FIG. 9: Some inflations of the Triangle structure.

4. Exemplary Inflations of The Triangle Structure

5. Deriving Symmetric Causal Compatibility Inequalities

Quite surprisingly, it is possible to derive causal compatibility inequalities for the Triangle structure that are symmetric under the exchange of parties. This is done by grouping marginal events $m \in \mathcal{S}(\mathcal{M})$ into orbits under the action of party exchange. Even more interestingly, these inequalities are violated by the Fritz distribution. This comes as a surprise because the Fritz distribution is not invariant under the exchange of parties; in fact the Fritz distribution is noticeably biased toward C.

In general, exploiting symmetries of the marginal scenario is useful for a few distinct reasons. First, Bancal *et al.* [49] discuss computational advantages in considering symmetric versions of marginal polytopes mentioned in Appendix A 3; the number of extremal points typically grows exponentially in $\mathcal{J}$, but only polynomial for the symmetric polytope. They also note a number of interesting inequalities (such as CHSH [19]) can be written in a way that is symmetric under the exchange of parties, demonstrating that non-trivial inequalities can be recovered from facets of a symmetric polytope. Secondly, numerical optimizations against symmetric inequalities lead to one of two interesting cases: either the extremal distribution is symmetric itself or it is not. The latter case generates a family of incompatible distributions obtained by applying symmetry operations on the extremal distribution[13].

To demonstrate what types of symmetries we have in mind, we present the following example:

Example 12. First consider the marginal scenario $\mathcal{M} = \{\{A, B, C\}, \{C, D\}, \{A, D\}\}$ where each variable in $\mathcal{J} = \{A, B, C, D\}$ has binary outcomes: $O_A = O_B = O_C = O_D = \{0, 1\}$. Moreover, let $I_{\mathcal{M}}$ be the following compatibility inequality over $\mathcal{M}$ for some causal structure $\mathcal{G}$,

$$I_{\mathcal{M}} = \{P_{ABC}(010) \leq P_{CD}(00) + P_{AD}(01)\}$$

Intuitively, the compatibility of $\mathcal{P}^{\mathcal{M}}$ is *independent* of the how one labels outcomes. Suppose $\varphi : \mathcal{J} \to \mathcal{J} : \{a, b, c, d\} \mapsto \varphi[\{a, b, c, d\}] = \{c, a, d, b\}$ is a permutation of the variables $\mathcal{J}$. Intuitively φ acts on $I_{\mathcal{M}}$ in the following way,

$$\begin{aligned}\varphi[I_{\mathcal{M}}] &= \left\{P_{\varphi[abc]}(010) \leq P_{\varphi[cd]}(00) + P_{\varphi[ad]}(01)\right\} \\ &= \{P_{cad}(010) \leq P_{db}(00) + P_{cb}(01)\} \\ &= \{P_{acd}(100) \leq P_{bd}(00) + P_{bc}(10)\}\end{aligned}$$

In general (as in this example), permutations over all variables can modify the marginal scenario $\varphi[\mathcal{M}] \neq \mathcal{M}$; in general we write $\varphi[I_{\mathcal{M}}] = \varphi[I]_{\varphi[\mathcal{M}]}$. Although $\varphi[I_{\mathcal{M}}]$ is still a valid causal compatibility inequality, in constrains marginal models of a different marginal scenario $\varphi[\mathcal{M}]$ and are in general of no interest.

Following the observations of Example 12, it is always possible to permute the random variables in a marginal scenario to produce new and valid inequalities. The symmetry group over a the complete set of variables $\mathcal{J}$ is simply the permutation group[14] $\mathsf{Perm}(\mathcal{J})$. As was noted in Example 12, some permutations $\varphi \in \mathsf{Perm}(\mathcal{J})$ have the undesired possibility of transforming $\mathcal{M}$ into a different marginal scenario.

$$\varphi[\mathcal{M}] \equiv \{\varphi[V_1], \ldots, \varphi[V_k]\} \neq \mathcal{M}$$

This means that there could exist $V \in \mathcal{M}$ such that $\varphi[V] \notin \mathcal{M}$. Permutations φ retaining this property are unable to generate compatibility inequalities over the same marginal scenario. Therefore, the desired set of symmetries is a subgroup that takes $\mathcal{M}$ to $\mathcal{M}$.

Definition 13. The **variable permutation group** Φ for a marginal scenario $\mathcal{M}$ is the joint permutation subgroup that stabilizes the marginal scenario.

$$\Phi(\mathcal{M}) \equiv \{\varphi \in \mathsf{Perm}(\mathcal{J}) \mid \forall V \in \mathcal{M} : \varphi[V] \in \mathcal{M}\}$$

Obtaining the variable permutation group $\Phi(\mathcal{M})$ is a simple application of a group stabilizer algorithm. After obtaining $\Phi(\mathcal{M})$, one can take known compatibility inequalities $I_{\mathcal{M}}$ can create a whole family of inequalities $\{\varphi[I_{\mathcal{M}}] \mid \varphi \in \Phi(\mathcal{M})\}$ that are valid for the same marginal scenario $\varphi[I_{\mathcal{M}}] = \varphi[I]_{\varphi[\mathcal{M}]} = \varphi[I]_{\mathcal{M}}$.

Although very useful for reducing computational complexity [49], we divert our attention to finding *symmetric* inequalities themselves.

$$\forall \varphi \in \Phi(\mathcal{M}) : \varphi[I_{\mathcal{M}}] = I_{\mathcal{M}}$$

In order to do this, we must first define how elements $\varphi \in \Phi$ act on marginal events. The action of $\varphi \in \Phi$ on any outcome $f \in \mathcal{S}(V)$ (denoted $\omega[f]$) is defined as,

$$\forall v \in \mathcal{D}(\varphi[f]) : (\varphi[f])(v) \equiv f\left(\varphi^{-1}[v]\right) \tag{A13}$$

Equation (A13) is nothing more than a mathematical definition for the intuitive action used in Example 12. Through repeated action of $\varphi \in \Phi$ on marginal outcomes $m \in \mathcal{S}(\mathcal{M})$ and joint outcomes $j \in \mathcal{S}(\mathcal{J})$, one can define **group orbits** of Φ in $\mathcal{S}(\mathcal{M})$ and $\mathcal{S}(\mathcal{J})$.

$$\begin{aligned}\mathsf{Orb}_{\Phi}(m) &\equiv \{\varphi[m] \mid \varphi \in \Phi\} \\ \mathsf{Orb}_{\Phi}(j) &\equiv \{\varphi[j] \mid \varphi \in \Phi\}\end{aligned}$$

Using these group orbits, it is possible to contract the incidence matrix M of a marginal scenario into a symmetrized version.

---

[13] If the extremal distribution happens to not be symmetric, there is indication that the space of accessible distributions is non-convex.
[14] The permutation group over a set S is the set of all bijective maps $\varphi : S \to S$.

Definition 14. The **symmetric incidence matrix** M_Φ for a marginal scenario $\mathcal{M}$ and the variable permutation group Φ is a contracted version of the incidence matrix M for $\mathcal{M}$. Each row of M_Φ corresponds to a marginal orbit $\mathsf{Orb}_\Phi(m)$. Analogously each column of M_Φ corresponds to a joint orbit $\mathsf{Orb}_\Phi(j)$. The entries of M_Φ are integers and correspond to summing over the rows and columns of M that belong to each orbit.

$$(M_\Phi)_{\mathsf{Orb}_\Phi(m),\mathsf{Orb}_\Phi(j)} = \sum_{\substack{j'\in\mathsf{Orb}_\Phi(j)\\ m'\in\mathsf{Orb}_\Phi(m)}} M_{m',j'}$$

It is possible to analogously define a **symmetric joint distribution vector** $\mathcal{P}_\Phi^{\mathcal{J}}$ indexed by $\mathsf{Orb}_\Phi(j)$,

$$\left(\mathcal{P}_\Phi^{\mathcal{J}}\right)_{\mathsf{Orb}_\Phi(j)} = \sum_{j'\in\mathsf{Orb}_\Phi(j)} \mathcal{P}_j^{\mathcal{J}}$$

And a **symmetric marginal distribution vector** $\mathcal{P}_\Phi^{\mathcal{M}}$ indexed by $\mathsf{Orb}_\Phi(m)$,

$$\left(\mathcal{P}_\Phi^{\mathcal{M}}\right)_{\mathsf{Orb}_\Phi(m)} = \sum_{m'\in\mathsf{Orb}_\Phi(m)} \mathcal{P}_m^{\mathcal{M}}$$

Which altogether define a new, **symmetric marginal problem**,

$$\mathcal{P}_\Phi^{\mathcal{M}} = M_\Phi \cdot \mathcal{P}_\Phi^{\mathcal{J}}$$

The symmetric marginal problem can be solved using the same computational methods discussed in Appendix A 3 and will produce symmetric inequalities.

In the context of the Inflation technique, a variable symmetry $\Phi(\mathcal{M}')$ over an inflated marginal scenario $\mathcal{M}'$ does not always correspond to a variable symmetry under deflation $\Phi(\mathcal{M})$. In order to derive deflated inequalities that are symmetric under an exchange of parties, it is required that $\Phi(\mathcal{M}') \sim \Phi(\mathcal{M})$ are equivalent up to copy-index.

For the Triangle structure in particular, the variable permutation group is simply the set of permutations on A, B, C: $\Phi(\mathcal{M}) = \mathsf{Perm}(A, B, C)$. For the Web inflation (Fig. 9d), we have obtained $\Phi\big(\mathsf{PreInj}_{\mathcal{G}}(\mathcal{G}')\big)$ which is an order 48 group with the following 4 generators:

$$\begin{array}{cccc}
\varphi_1 & \varphi_2 & \varphi_3 & \varphi_4 \\
A_1 \to A_4 & A_1 \to A_1 & A_1 \to C_1 & A_1 \to A_1 \\
A_2 \to A_3 & A_2 \to A_3 & A_2 \to C_2 & A_2 \to A_2 \\
A_3 \to A_2 & A_3 \to A_2 & A_3 \to C_3 & A_3 \to A_3 \\
A_4 \to A_1 & A_4 \to A_4 & A_4 \to C_4 & A_4 \to A_4 \\
B_1 \to B_4 & B_1 \to C_1 & B_1 \to A_1 & B_1 \to B_2 \\
B_2 \to B_3 & B_2 \to C_3 & B_2 \to A_2 & B_2 \to B_1 \\
B_3 \to B_2 & B_3 \to C_2 & B_3 \to A_3 & B_3 \to B_4 \\
B_4 \to B_1 & B_4 \to C_4 & B_4 \to A_4 & B_4 \to B_3 \\
C_1 \to C_4 & C_1 \to B_1 & C_1 \to B_1 & C_1 \to C_3 \\
C_2 \to C_3 & C_2 \to B_3 & C_2 \to B_2 & C_2 \to C_4 \\
C_3 \to C_2 & C_3 \to B_2 & C_3 \to B_3 & C_3 \to C_1 \\
C_4 \to C_1 & C_4 \to B_4 & C_4 \to B_4 & C_4 \to C_2
\end{array}$$

It is easy to verify that $\varphi_1, \varphi_2, \varphi_3, \varphi_4$ are all automorphisms of the web inflation. Moreover they stabilize the pre-injectable sets of Eq. (13). Importantly,

$$\Phi\big(\mathsf{PreInj}_{\mathcal{G}}(\mathcal{G}')\big) \sim \mathsf{Perm}(A, B, C)$$

To see this, φ_1 and φ_4 become the identity element $\mathbb{1}$ in $\mathsf{Perm}(A, B, C)$ upon removal of the copy-index, leaving φ_2 to generate reflections and φ_3 to generate rotations.

The symmetric incidence matrix M_Φ for the Web inflation is considerably smaller than M. The number of rows of M_Φ is a number of distinct orbits $\mathsf{Orb}_\Phi(m)$ in $\mathcal{S}(\mathcal{M})$. Likewise the number of columns is the number of distinct orbits $\mathsf{Orb}_\Phi(j)$ in $\mathcal{S}(\mathcal{J})$. For the Web inflation in particular, M_Φ is $450 \times 358,120$. Using the symmetric incidence matrix and linear programing methods, an infeasibility certificate was found that is capable of witnessing the Fritz distribution. This inequality is reproduced in Eq. (15).

[1] P. W. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer," SIAM J. Comp. **26**, 1484 (1997).

[2] S. Jordan, "Quantum algorithm zoo," (2016).

[3] C. H. Bennett and G. Brassard, "Quantum cryptography: Public key distribution and coin tossing," Theor. Comp. Sci. **560**, 7 (2014).

[4] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2011) p. 109.

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

[6] A. Einstein, B. Podolsky, and N. Rosen, "Can quantum-mechanical description of physical reality be considered complete?" Phys. Rev. **47**, 777 (1935).

[7] 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 (2012), arXiv:1208.4119.

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

[9] J. Pearl, "Causal inference in statistics: An overview," Stat. Surv. **3**, 96 (2009).

[10] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, "Bell nonlocality," Rev. Mod. Phys **86**, 419 (2013), arXiv:1303.2849.

[11] T. Fritz, "Beyond Bell's Theorem: Correlation Scenarios," New J. Phys **14**, 103001 (2012), arXiv:1206.5115.

[12] T. Fritz, "Beyond Bell's Theorem II: Scenarios with arbitrary causal structure," Comm. Math. Phys. **341**, 391 (2014), arXiv:1404.4812.

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

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

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

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

[17] E. Wolfe, R. W. Spekkens, and T. Fritz, "The inflation technique for causal inference with latent variables," (2016), arXiv:1609.00672.

[18] B. M. Terhal and D. P. DiVincenzo, "Adaptive quantum computation, constant depth quantum circuits and arthur-merlin games," Quant. Inf. Comp. **4**, 134 (2004).

[19] 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).

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

[21] D. Petz and L. Ruppert, "Optimal quantum-state tomography with known parameters," J. Phys. A **45**, 085306 (2015).

[22] S. R. Hedemann, "Hyperspherical parameterization of unitary matrices," (2013), arXiv:1303.5904.

[23] K. Fujii, K. Funahashi, and T. Kobayashi, "Jarlskog's parametrization of unitary matrices and qudit theory," Int. J. Geom. Meth. Mod. Phys. **03**, 269 (2006).

[24] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, "Measurement of qubits," Phys. Rev. A **64**, 052312 (2001).

[25] M. Grasmair, "On the existence of a cholesky factorization," (2014).

[26] C. Spengler, M. Huber, and B. C. Hiesmayr, "A composite parameterization of unitary groups, density matrices and subspaces," J. Phys. A **43**, 385306 (2010).

[27] C. Moler and C. V. Loan, "Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later," SIAM Rev. **45**, 3 (2003).

[28] K. F. Pál and T. Vértesi, "Maximal violation of a bipartite three-setting, two-outcome bell inequality using infinite-dimensional quantum systems," Phys. Rev. A **82**, 022116 (2010).

[29] J. Nocedal and S. Wright, *Numerical Optimization (Springer Series in Operations Research and Financial Engineering)* (Springer, 2000).

[30] D. J. Wales and J. P. K. Doye, "Global optimization by basin-hopping and the lowest energy structures of lennard-jones clusters containing up to 110 atoms," J. Phys. Chem. A **101**, 5111 (1997).

[31] A. A. Methot and V. Scarani, "An anomaly of non-locality," (2007), arXiv:quant-ph/0601210.

[32] T. Fritz and R. Chaves, "Entropic inequalities and marginal problems," IEEE Trans. Info. Theor. **59**, 803 (2011).

[33] N. N. Vorob'ev, "Consistent families of measures and their extensions," Theor. Prob. Applic. **7**, 147 (1962).

[34] S. Abramsky and A. Brandenburger, "The sheaf-theoretic structure of non-locality and contextuality," New J. Phys **13**, 113036 (2011), arXiv:1102.0264.

[35] S. Abramsky and L. Hardy, "Logical bell inequalities," Phys. Rev. A **85**, 062114 (2012).

[36] S. Mansfield and T. Fritz, "Hardy's non-locality paradox and possibilistic conditions for non-locality," Found. Phys. **42**, 709 (2011).

[37] I. Pitowsky, "Correlation polytopes: Their geometry and complexity," Math. Prog. **50**, 395 (1991).

[38] G. B. Dantzig and B. C. Eaves, "Fourier-motzkin elimination and its dual," J. Combin. Theor. A **14**, 288 (1973).

[39] 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).

[40] A. Schrijver, *Theory of Linear and Integer Programming* (Wiley, 1998).

[41] S. Lahaie, "How to take the dual of a linear program," (2008).

[42] Bradley, Hax, and Magnanti, *Applied Mathematical Programming* (Addison-Wesley, 1977) Chap. 4, pp. 143–144.

[43] E. D. Andersen, Comp. Optim. Applic. **20**, 171 (2001).

[44] MOSEK Optimization Suite, "Mosek IEEE Trans. Info. Theor." (2016).

[45] Gurobi Optimizer, "Farkasdual," (2016).

[46] IBM - CPLEX Optimization Studio, "Cplex," (2016).

[47] CVX - Matlab Software for Disciplined Convex Programming, "Dual variables," (2016).

[48] Martin Andersen, Joachim Dahl, and Lieven Vandenberghe, "Linear cone programs," (2016).

[49] J.-D. Bancal, N. Gisin, and S. Pironio, "Looking for symmetric bell inequalities," J. Phys. A **43**, 385303 (2010).