Quantum violation of Causal Friendliness in the instrumental scenario

Emanuele Polino,[1,*] Giulia Lelli,[2] Aaditya Datar,[1] Marwan Haddara,[3] Davide Poderini,[4] Giorgio Milani,[2] Sergei Slussarenko,[1] Howard M. Wiseman,[1] Nora Tischler,[1] Eric G. Cavalcanti,[3,†] and Fabio Sciarrino[2,‡]

[1] *Queensland Quantum and Advanced Technologies Research Institute, Griffith University, Yuggera Country, Brisbane, QLD 4111, Australia*
[2] *Dipartimento di Fisica - Sapienza Università di Roma, P.le Aldo Moro 5, I-00185 Roma, Italy*
[3] *Queensland Quantum and Advanced Technologies Research Institute, Griffith University, Yugambeh Country, Gold Coast, QLD 4222, Australia*
[4] *Università degli Studi di Pavia, Dipartimento di Fisica, QUIT Group, via Bassi 6, 27100 Pavia, Italy*

arXiv:2611.17431v1 [quant-ph] 12 Nov 2026

How far can quantum theory depart from a classical worldview? Bell inequality violations show that no theory can simultaneously uphold absoluteness of observed events, freedom of choice, relativistic causality, and classical causal explanations. While generalized notions of causality, capable of incorporating quantum resources, can account for Bell correlations, the Local Friendliness theorem showed that, in extended Wigner's friend scenarios, the conflict between quantum theory and causal explanations persists even after classical causality is abandoned. This revealed a new form of nonclassicality stronger than what was revealed by Bell's theorem. However, Local Friendliness has been confined to Bell causal structures so far, and it remains unknown whether a related form of nonclassicality persists in more general causal networks. Here we answer this question by considering a Wigner's friend version of the instrumental network, a paradigmatic causal structure in which communication between parties is allowed, and, unlike the Bell scenario, relativistic causality does not impose any restriction. We show, both theoretically and in a proof-of-principle photonic experiment, that quantum correlations are incompatible with Causal Friendliness, which is the conjunction of the absoluteness of observed events and the generalized nonclassical causal constraints imposed by the network. Such constraints extend relativistic causality from Bell scenarios to networks of arbitrary structure. These results establish a form of quantum nonclassicality stronger than previously observed in instrumental-network violations of classical causality, generalizing Local Friendliness beyond Bell-Wigner scenarios and uncovering a new class of correlations with potential applications in future quantum technologies.

I. INTRODUCTION

A central goal of science is not only to predict observations but also to explain them in terms of underlying causal mechanisms. From physics to medicine, causal explanations provide the conceptual framework through which observed correlations are understood [1]. Quantum theory, however, has repeatedly challenged this programme. Bell's theorem [2, 3] showed that no classical causal explanation can reproduce the observed correlations between distant quantum systems in the Bell causal structure, revealing a profound incompatibility between quantum theory and the classical causal worldview [4, 5]. This incompatibility, conventionally described as quantum nonlocality [3], invokes the classical notion of causes. In response to Bell's theorem, generalized causal frameworks have been proposed, allowing causal influences and common causes to be nonclassical, thereby accounting for correlations that elude classical causal models without abandoning the causal structure itself [4, 6–12].

More recently, however, the Local Friendliness (LF) theorem [13] has shown that abandoning classical causality does not resolve the tension between quantum theory and causal explanation [14, 15]. By combining the Bell scenario with a "Wigner's friend" and an assumption of Absoluteness of Observed Events (AOE), it demonstrates that quantum predictions are incompatible even with generalized causal explanations [13–15].

These developments leave open a fundamental question. Previous theoretical and experimental demonstrations [13, 15–18] of this form of nonclassicality based on AOE and generalized causal constraints have been formulated within Bell causal structures, where distant observers perform measurements on systems distributed by a common source. Indeed, the original LF theorem used an assumption based on the relativistic causal arrow principle [14] that does not impose constraints on scenarios without spacelike separation. While those constraints can be motivated from generalized causality in a Bell causal structure [15], it remains unknown whether they reflect a broader limitation of nonclassical causal explanations in general causal networks.

Here, we show that indeed this form of nonclassicality is not confined to Bell causal structures. We do this by considering the instrumental scenario, a paradigmatic causal network with a central role in both causal inference and quantum networks [1, 19–23]. In the instrumental network, the outcome of one observer, Alice, can directly influence the measurement performed by another observer, Bob (see Fig. 1), so spacelike separation plays no role and the relativistic causal arrow assumption of the Local Friendliness theorem no longer pro-

* e.polino@griffith.edu.au , emanuele.polino@gmail.com
† e.cavalcanti@griffith.edu.au
‡ fabio.sciarrino@uniroma1.it

vides the relevant constraint. We introduce a Wigner's-friend extension of the instrumental scenario and formulate a set of theory-independent assumptions that we call *Causal Friendliness* (CF). These assumptions combine AOE with constraints derived for the instrumental causal structure using generalized causal models [8] based on Operational Probabilistic Theories (OPTs). This framework extends the description of causal relations beyond classical and quantum theory while remaining compatible with the no-signaling principle [8, 24].

A violation of CF, formalised in Sec. II B, reveals a strong form of nonclassicality beyond the failure of classical causal models (CCMs), ruling out causal explanations even when the underlying resources and influences are allowed to be nonclassical. We prove that quantum theory violates CF in the instrumental scenario (Sec. III) and demonstrate this violation in a proof-of-principle photonic experiment using polarization-entangled photons, active feed-forward, and a single-photon path to encode the friend's outcome (Sec. IV). The observed correlations violate a device-independent Bell-like CF inequality by more than 120 standard deviations, without relying on a characterization of the internal operation of the experimental devices. These results reveal the tension between quantum theory and models that respect generalized causality in the assumed causal structure and the AOE. By taking this strong form of quantum nonclassicality beyond the Bell scenario and into a broader landscape of causal structures, our work opens new avenues for quantum foundations and quantum technologies.

II. CAUSAL FRIENDLINESS BEYOND THE BELL SCENARIO

A. Friendliness: nonclassicality beyond the violation of classical causality

Bell's theorem is traditionally understood as ruling out local hidden-variable theories [2, 3]. From the modern perspective of causal inference [4, 6] however, it is viewed as the incompatibility between quantum correlations and four basic assumptions, which, following Refs. [5, 14], are: (i) Absoluteness of Observed Events (AOE), (ii) relativistic causal arrow, (iii) freedom of choice (or independent interventions [14]), and (iv) classical causal explanation. In particular, classical causal explanations are a much debated concept in physics [4, 5, 25, 26]. In the Bell scenario, the classical causal notion is embodied in *Reichenbach's principle of common cause* [7, 14]: that for any two sets of correlated events **A** and **B**, such that no event in either is a cause of any event in the other, there exists a set of common causes **C**, such that conditioning on **C** eliminates the correlation. Reichenbach's principle follows from the more general *classical causal Markov condition* [1]: that a set of events is independent of any non-effects given a sufficient specification of its direct causes. Violations of the classical causal Markov condition have been studied [22, 27–31] and experimentally demonstrated [21, 32–38] in various causal networks beyond the Bell scenario. These observations have motivated a broad research programme aimed at extending classical causal explanations to the quantum, or more generally nonclassical (OPT) regime, to explain quantum nonclassicality [4, 6, 8–12].

However, existing frameworks of generalized causal explanations are challenged by the quantum measurement problem—the tension between the deterministic unitary evolution of quantum theory and the emergence of definite measurement outcomes [39–47] best illustrated by Wigner's friend scenarios [48–51]. When extended to quantum systems distributed among distant parties [45, 52, 53], these scenarios reveal that even generalized causal explanations, which correspond to relativistic causality in Bell scenarios, become incompatible with quantum theory if AOE and free choices are retained [13–15, 17, 18, 54]. We stress that, although the LF assumptions were originally introduced in the context of extended Wigner's friend experiments, which provide a natural motivation for AOE, such scenarios represent only one possible perspective in which the CF assumptions may be considered. More generally, CF can be viewed as a theory-independent relaxation of classical causal constraints defining partially deterministic models, which also have application to quantum information protocols such as device-independent randomness certification [55].

We now extend this framework beyond Bell scenarios, to the instrumental scenario, which therefore requires replacing Bell-specific relativistic causal arrow constraints with generalized causal constraints.

B. Causal Friendliness in the Instrumental Scenario

The (classical) instrumental causal structure, shown in Fig. 1(a), consists of a latent classical common cause Λ acting on A and B, a direct influence from A's outcome to B, and an "instrumental" variable (here, the measurement setting) X that controls A. It can be viewed as a relaxed Bell scenario in which one-way communication of outcomes from A to B is allowed, and only A is influenced by a free setting. Instrumental tests are central to causal inference because they enable us to distinguish between the two possible classical explanations for the correlations between A and B: a common cause or a direct cause [1, 19, 20]. For this reason, this scenario is widely applied across scientific disciplines [56, 57].

In the classical causal framework, the Markov condition [1] applied to the instrumental network imposes the following constraint on the probabilities: $P(a,b|x) = \sum_\lambda P(a|x,\lambda)\,P(b|a,\lambda)\,P(\lambda)$. This constraint leads to inequalities [58] that can be violated by quantum systems, as shown theoretically and experimentally in Ref. [21]. A connection between violations of standard Bell and

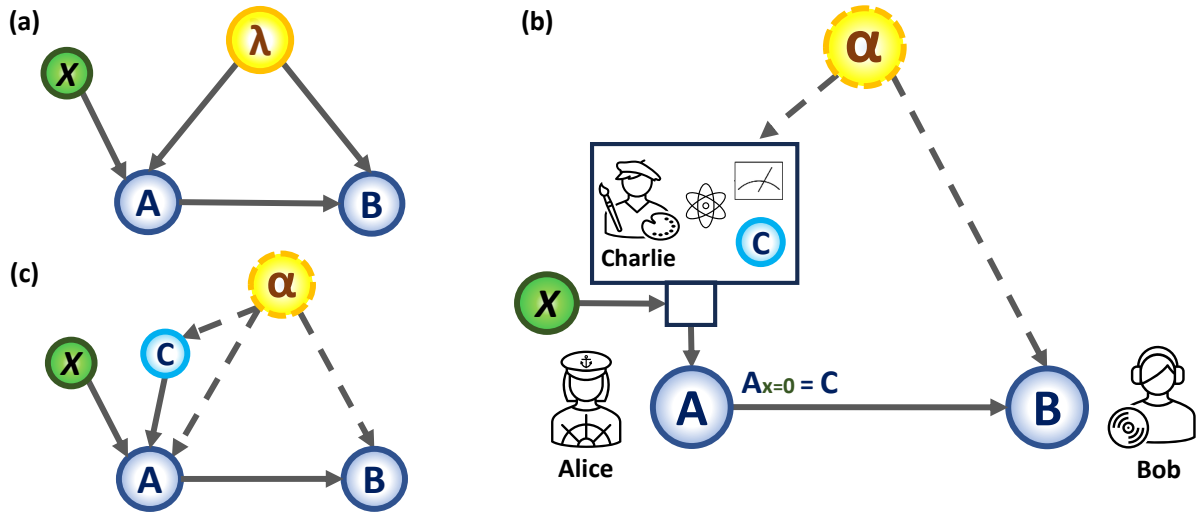


FIG. 1: **(a)** Standard classical causal structure of the instrumental scenario, where a single (classical) source λ shares correlations between the two measurement nodes A and B, representing the observed variables of the network. The variable A (Alice's measurement), influenced by the free choice X, can affect Bob's measurement variable B. **(b)** Extended Wigner's friend scenario built within the instrumental framework, where an outcome C is produced after the interaction between the incoming subsystem and an observer system in an isolated laboratory. The laboratory is then measured by a superobserver with a measurement choice X and outcome A. The assumptions in Eqs. (1)-(2) imply that, whenever $X = 0$, $A = C$. **(c)** Causal diagram of the variables involved in the Instrumental Wigner's friend scenario, where the presence of two channels from the (nonclassical) source α to the node A is made explicit: one classical channel (acting when $X = 0$) and one nonclassical channel (acting when $X \neq 0$). The dashed (solid) arrows represent the causal influences from a nonclassical (classical) node.

instrumental inequalities has also been established [22]. The instrumental scenario has further been explored in the contexts of post-quantum steering [59, 60], as well as forms of nonclassicality enabled by interventions, i.e., actions assumed to be able to force the variable A to assume certain values independently of its causal parents [23, 61].

Now consider Fig. 1(b), illustrating a generalization of the instrumental scenario to one involving Wigner's friend. The subsystem in Alice's station is first measured by her friend Charlie inside an isolated laboratory. Alice acts as a superobserver outside the laboratory and can perform quantum measurements on the joint system composed of Charlie and the subsystem. Charlie always performs a fixed measurement on the subsystem, yielding an outcome denoted by C. For one of Alice's measurement settings, say $X = 0$, she simply opens the laboratory and asks Charlie for the value of C, and records $A = C$. For all other values of X, Alice performs a different measurement on the composite Charlie–subsystem laboratory. Note that, under the assumption of AOE (Absoluteness of Observed Events, stated rigorously below), C is always as real as A. However, this is not tantamount to assuming that Charlie's measurement collapses the quantum wavefunction; there are quantum interpretations where certain properties have well-defined values even when the quantum state is not in the associated eigenstate [62], and the no-go theorem has to apply to all conceivable theories satisfying the theorem's assumptions.

So far, we have described one station, which consists of one superobserver (Alice) and one friend (Charlie). There is another station, with a regular observer (Bob), but, unlike in Refs. [13, 14, 17], these stations are not subject to the usual bipartite Bell causal structure. Rather, in this instrumental scenario, after completing her measurement, Alice sends her outcome to Bob's station, where the second subsystem is measured in a way that may depend on Alice's result, and yields an outcome B. Following the framework of [15], the generalized causal structure corresponding to this scenario is depicted in Fig. 1(c): the subsystems held by the two stations share a *nonclassical* common cause α. The dashed arrows thus represent nonclassical channels while the solid ones are classical channels; X, C, A, and B are observed (classical) variables.

Given this scenario, we now list and explain the CF assumptions that constrain the observed probability distributions $\wp(a, b|x)$:

(1) Absoluteness of Observed Events (AOE). AOE imposes that each observed event has a value that is absolute, i.e., not relative to anything (for discussions on this assumption in extended Wigner's friend experiments, see Refs. [14, 17, 18, 51, 63, 64]). In our scenario, this implies the existence of a theoretical joint probability distribution $P(a, c, b|x)$ that reproduces the observed

probability $\wp(a,b|x)$:

$$\wp(a,b|x) = \sum_c P(a,c,b|x) \;\; \forall\, a,b,x\,. \tag{1}$$

The description of the scenario for the setting $X = 0$ also motivates the auxiliary *tracking assumption* [18], that when Alice queries Charlie's outcome, the value he reports corresponds to the value C that is assumed to exist by AOE:

$$P(a|c,x=0) = \delta_{a,c} \;\; \forall\, a,c\,. \tag{2}$$

An equivalent constraint to the perfect tracking in Eq. (2) can nevertheless be obtained from a weaker assumption. One may indeed allow imperfect correlations between A and C—for example, due to classical noise in the friend's measurement or report—without introducing any additional nonclassical correlations between A and B. To cover these more general situations, the tracking assumption may be formulated more broadly by requiring only that, when $X = 0$, the information communicated from Charlie to Alice is related to A through a classical channel [solid arrow between C and A in Fig. 1(c)], with no nonclassical side channel [dashed arrow between α and A in Fig. 1(c)]. We refer to this screening-off assumption as a *reliable friend*. Importantly, irrespective of the underlying description, whether based on the tracking assumption or on a classical-channel model of the reliable friend, Eq. (2) provides the mathematical constraint relevant to CF [65]. As shown in the following sections and discussed in Appendix H, the reliable-friend assumption can be further relaxed in a distinct way by allowing a limited nonclassical contribution to the channel from the source to A when $X = 0$. This weakens the constraint in Eq. (2) and, consequently, CF as a whole.

(2) Generalized causality constraints in the instrumental scenario. These constraints are derived from the OPT generalized Markov condition introduced by Ref. [8]. This condition is expressed in an abstract diagrammatic framework and represents a relaxation of classical causality, allowing nonclassical causal relations between variables in a theory-independent way, and extending the no-signaling condition to arbitrary nonclassical causal structures. We refer to Appendix A and Ref. [8] for more details.

We first consider the constraint arising from generalized causality applied to the instrumental causal structure, which reflects the absence of a directed causal path from X to C, and which is motivated by choosing X via suitable external variables. Commonly called *Freedom of Choice*, or *No-Superdeterminism*, it requires statistical independence between Alice's choice X and the variable C:

$$P(x|c) = P(x) \quad \forall\, x,c \;\; . \tag{3}$$

Next, we consider the constraints associated with the absence of a direct causal arrow from X to B. In the instrumental scenario, the OPT causality constraints allow for the possibility that X can signal to B, due to the existence of a causal path via A, so that $\wp(b|x) \neq \wp(b)$ in general. Nevertheless, the absence of a direct causal arrow from X to B, together with the freedom-of-choice constraint above, imposes the following inequality constraints (details in Appendix C) [8]:

$$\begin{gathered} P(a,b=-1|x,c) + P(a,b=+1|x',c) \leq 1 \\ \forall\; x \neq x',\;\; a\;, c\;. \end{gathered} \tag{4}$$

These inequalities are identical, except for the additional conditioning on c, to the standard *Pearl inequalities* [20], known to hold in both classical and OPT instrumental scenarios [8] (see Appendix C). We note that, for the particular case of the instrumental scenario, these conditions are mathematically equivalent to the constraints arising from the exclusivity principle [66, 67] used in the graph-theoretic approach to noncontextual, local hidden variable models, and also in networks with hybrid assumptions [68] (see Appendix D). Together, Eqs. (1)–(4) define Causal Friendliness [69].

III. QUANTUM VIOLATION OF CAUSAL FRIENDLINESS IN THE INSTRUMENTAL SCENARIO

The assumptions in Eqs. (1)–(4) defining CF give rise to partially deterministic behaviours [13, 70], whereby deterministic outcomes occur only for a subset of inputs (here, $X = 0$). For all the remaining inputs, the outcome statistics are constrained solely by the generalized causality principle. We consider the minimal scenario where a quantum-classical gap can be found [21, 22]: dichotomic outputs $A, B \in \{-1, +1\}$, and three choices of measurements $X \in \{0, 1, 2\}$. Equations (1)-(2) imply the output A is deterministic when the measurement choice is $X = 0$. We fully characterize the resulting CF correlations by finding the polytope resulting from the constraints (1)–(4). In general, the following strict inclusion relations between various sets of correlations hold: $CCM \subset CF \subset OPT$, and $CCM \subset Q \subset OPT$. Here, the sets correspond to classical causal models CCM, Causal Friendliness CF, quantum Q, and operational probabilistic theories OPT. Note that the relation between the quantum set Q and the Causal Friendliness set CF is not one of inclusion; the two are distinct, intersecting sets. The CF instrumental polytope has 36 vertices and 48 facets. Among these facets, 8 are genuine CF, in the sense that they are not facets of the CCM nor of the OPT polytopes.

The 8 genuine CF facets, up to relabeling of the outcome values of the variables A, B and the two nondeterministic input values of X in the observed probability $\wp(a,b|x)$, are described by a single class of genuine CF inequalities:

$$I_{\text{inst}} \equiv \langle AB\rangle_1 - \langle AB\rangle_0 + \langle B\rangle_0 - \langle B\rangle_2 \overset{CF}{\leq} 2\;. \tag{5}$$

The recorded outcomes are binary outcomes taking values $a, b \in \{-1, +1\}$, and the sufficient statistics are the mean of Bob's outcome, $\langle B\rangle_x = \sum_{a,b} b \cdot \wp(a, b|x)$, and the mean product of the outcomes, $\langle AB\rangle_x = \sum_{a,b} a \cdot b \cdot \wp(a, b|x)$.

Remarkably, quantum systems using two-qubit entangled states and projective measurements can violate Eq. (5), showing that Causal Friendliness can be disproven in the instrumental scenario. To see this, consider the maximally entangled state: $|\Psi_{AB}\rangle = (|01\rangle - |10\rangle)/\sqrt{2}$. Alice performs a binary projective measurement specified by an observable $\hat{A}_x$, while Bob measures an observable $\hat{B}_y$, with outcomes $a, b \in \{-1, +1\}$. Bob's setting is determined by Alice's outcome, and conventionally we set $y = a$. The corresponding outcome projectors are $\Pi^A_{a|x} = (\mathbb{I}_A + a \cdot \hat{A}_x)/2$, $\Pi^B_{b|y} = (\mathbb{I}_B + b \cdot \hat{B}_y)/2$. The adaptive protocol therefore has joint measurement effects:

$$E_{ab|x} = \Pi^A_{a|x} \otimes \Pi^B_{b|a}, \tag{6}$$

yielding probabilities $\wp(a, b|x) = \langle \Psi_{AB}|E_{ab|x}|\Psi_{AB}\rangle$. For each x, these effects satisfy $\sum_{a,b} E_{ab|x} = \mathbb{I}_{AB}$. Choosing observables

$$\hat{A}_0 = -\sigma_z, \ \hat{A}_1 = (\sigma_z - \sigma_x)/\sqrt{2}, \ \ \hat{A}_2 = (\sigma_z + \sigma_x)/\sqrt{2}$$

for Alice, and

$$\hat{B}_{+1} = \sigma_x, \ \hat{B}_{-1} = -\sigma_z$$

for Bob, yields $I_{\text{inst}} = 1 + \sqrt{2} > 2$, thereby violating the CF bound. We also prove that this is an optimal choice by showing that the achieved value for the violation matches the numerical upper bound obtained at the second level of the Navascués–Pironio–Acín hierarchy [71]. The OPT bound for Eq. (5) is instead 3.

These findings establish a central result of this work:

Theorem 1. *Quantum predictions in the instrumental scenario are incompatible with constraints implied by Causal Friendliness.*

In the Wigner's friend narrative, to observe a quantum violation of Causal Friendliness, the superobserver must coherently control both the friend and the relevant environment, while the friend's role as a genuine observer must remain well justified [17]. We note, however, that the Wigner's friend narrative represents only a framework for CF, and the theorem and the inequalities found are fully theory-independent and "narrative"-independent, since they follow from the mathematical assumptions in Eqs. (1)-(4). These represent a strict relaxation of the standard CCM assumptions, and thus CF violation implies a notion of nonclassicality stronger than the violation of classical causal models.

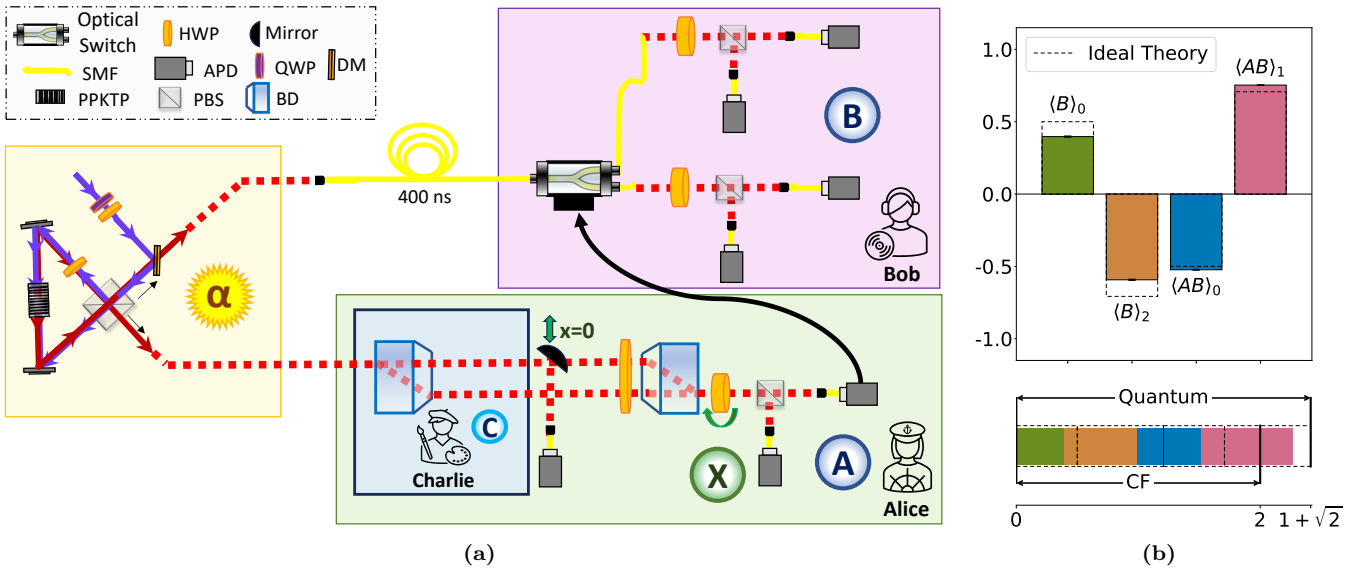
It is interesting to note that we have quantum violations of genuine CF inequalities with just three measurement settings, i.e., $|X| = 3$, when A and B are dichotomic. That is, no additional measurement settings are required to violate the assumptions of CF beyond those needed to violate standard classical causal constraints. Note that we are not considering scenarios where one can use interventions that would allow one to violate classical causal constraints with only two measurements [23, 61]. This differs from the standard Bell scenario, where demonstrating a genuine violation of LF necessitates a larger set of measurement inputs than those sufficient to violate local causality [13].

IV. EXPERIMENTAL DEMONSTRATION

We perform a photonic experiment implementing the Wigner's friend instrumental scenario to test, through quantum correlations, the constraints imposed by CF. We generate polarization-entangled photon pairs from a spontaneous parametric down-conversion (SPDC) source based on a Sagnac interferometer geometry [72]. One of the photons is sent to Alice's station, where a coupling between the path and the polarization is realized through a beam displacer (see Fig. 2(a)). This coupling, in the Wigner's friend narrative, represents the (reversible) measurement of the system, encoded in the photon's polarization, by the friend, encoded in the photon's path. When the choice of the superobserver is $X = 0$, a motorized mirror is inserted, and the path of the photon is directly measured by single-photon detectors in the two paths coupled with the horizontal or vertical polarization, respectively. If the measurement choice is $X \neq 0$, the path–polarization coupling is reversed by a second beam displacer, followed by the implementation of a polarization measurement. The second photon is first delayed by about 400 ns and then sent to Bob's station. There, the photon enters a fast optical switch that routes it to one of two measurement stations conditioned on Alice's outcome: a click in the $+1$ detector triggers an electronic feed-forward signal that changes the selected output port of Bob's photon, whereas a click in the -1 detector leaves the switch in its default configuration. After each of the two output ports of the switch, a static polarization measurement is performed followed by single-photon detection. Through this feed-forward, the output bit from Alice's measurement determines Bob's measurement. (See Appendix G for experimental details and the outcome assignment strategy.)

We begin by testing the genuine CF facet in Eq. (5), and we measure $I^{\text{exp}}_{\text{inst}} = 2.266 \pm 0.002$, which exceeds the CF bound by more than 120 standard deviations (see Fig. 2(b)). The uncertainty is evaluated as the standard error of the mean over 100 repetitions of the measurement, which automatically captures both Poissonian counting statistics and possible additional fluctuations due to experimental instabilities.

Furthermore, we engineer a noisier distribution obtained by collecting the data with an enlarged coincidence window, causing an increase in accidental counts



(a) (b)

FIG. 2: **Experimental setup and results. (a)** A polarization-entangled photon source generates the singlet state via a type-II SPDC process in a Sagnac interferometer. One photon is sent to Alice's station, where it undergoes the friend interaction and the final superobserver's measurements, whose settings are labeled by X. The other photon goes through a delay line and is then directed to one of Bob's measurement stations by an optical switch using a feed-forward signal based on Alice's measurement outcome. APD = avalanche photodiode single-photon detector; BD = beam displacer; DM = dichroic mirror; HWP = half-wave plate; PBS = polarizing beam splitter; PPKTP = periodically poled potassium titanyl phosphate; QWP = quarter-wave plate; SMF = single-mode fiber. **(b)** Experimental correlator values estimated from the experimental statistics that violate the genuine CF facet in Eq. (5). The error bars (barely visible) represent 3σ, where the standard error σ on the mean values is calculated by repeating the measurement 100 times. The dashed grey lines indicate the values of the correlators calculated from an ideal maximally entangled singlet state and optimal measurements. The linear combination of the correlators generates the violation of the CF bound as depicted in the bottom part of the figure.

and acting substantially as a white noise on the correlations. This regime is particularly significant: although the resulting correlations remain nonclassical—violating an instrumental CCM facet—they lie entirely within the CF polytope. We observe a violation of a CCM facet by more than 39 standard deviations, while all 48 CF facets are satisfied within at least 13 standard deviations. This provides direct experimental evidence for the existence of correlations that are incompatible with classical causality but compatible with CF. The locations of both the experimental points in a two-dimensional projection of the OPT, CF, quantum, and CCM correlation sets are shown in Fig. 3.

Using the framework introduced in Ref. [73], we perform an additional analysis where the constraint in Eq. (2) is relaxed by allowing non-deterministic outcomes even for the setting $X = 0$. Specifically, we can introduce an ϵ-relaxation of Eq. (2) with the weaker condition $P(a = c|x = 0, c) \geq 1 - \epsilon \quad \forall c$. At first sight, the ϵ-relaxation could be interpreted as a trivial tracking failure of the friend, where the friend's record of the variable C is noisy in a classical sense. However, as discussed in Sec. II B, such classical noise alone does not affect the CF model. Rather, we interpret this relaxation as allowing the most general model, compatible with the ϵ-relaxed condition, for reproducing the observed correlations. With respect to the OPT causal model in Fig. 1(c), the ϵ-relaxation corresponds to permitting the nonclassical channel (the direct nonclassical arrow from the nonclassical source to node A) to remain active even when $X = 0$ in some fraction of the trials, rather than being completely switched off by the friend's classical channel. By characterizing the corresponding enlarged polytope, we find that the observed violation of Eq. (5) in our experiment also excludes models with an ϵ-relaxation up to $\epsilon = 13.1\%$ compatible with the observed violation within two standard deviations (see Appendix H for details).

V. DISCUSSION

In the last few decades, extensions of Bell's theorem in two distinct directions have reshaped our understanding of the relation between quantum phenomena and causality. One extends quantum violations of classical causal models beyond Bell scenarios to networks [8, 27, 29], including those where communication is allowed and spacelike separation is unnecessary or impossible, such as in

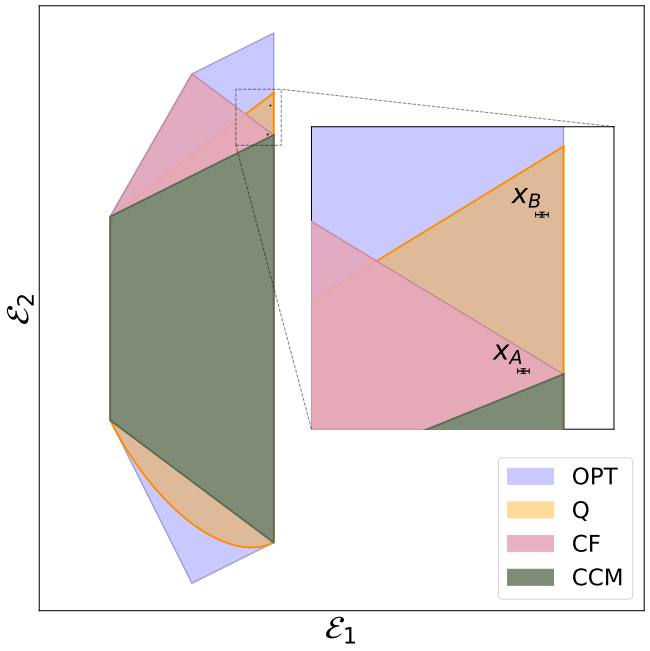


FIG. 3: **Correlation sets.** 2-D projection of correlation sets corresponding to the Operational Probabilistic Theories (OPT), quantum theory (Q), Causal Friendliness (CF), and Classical Causal Models (CCM). The experimental data are plotted as points to visualize the violations they imply: the point X_B lies outside both the CF and CCM polytopes, whereas the point X_A, obtained from noisier measurements, lies entirely within the CF polytope but outside the CCM polytope. The error bars represent 3σ, where σ is the standard error of the mean value in 100 repetitions. The expressions $\mathcal{E}_1$ and $\mathcal{E}_2$ on the axes are linear combinations of specific correlations; further details are provided in Appendix E.

the instrumental scenario [21]. The other, developed through extended Wigner's friend scenarios, shows that quantum predictions can remain incompatible even with generalized nonclassical causal explanations when Absoluteness of Observed Events is retained [13–15], but has so far been restricted to Bell causal structures, with a single source connecting two or more distant parties. In this work, we combined these two lines of research, revealing that this stronger form of quantum nonclassicality persists in causal networks beyond the Bell scenario.

We introduced Causal Friendliness in a Wigner's friend extension of the instrumental network, and used a generalized nonclassical causal framework to study this scenario. We proved that the corresponding quantum correlations are incompatible with any model that jointly satisfies AOE and the generalized OPT causal constraints associated with the instrumental structure. The significance of this result is that it extends the reach of Local-Friendliness-type no-go theorems and the consequent nonclassicality to causal networks beyond Bell scenarios. In particular, it shows that the tension between quantum predictions and the assumptions of AOE and generalized causality appears in scenarios where not only relativistic causal constraints impose no restriction, but which indeed cannot be realized with spacelike separation (which can be an intricate notion in contexts where AOE is under scrutiny [63]).

We further realized a proof-of-principle photonic experiment, in which the friend is encoded in the path degree of freedom of a single photon. The observed violation of the corresponding genuine CF inequality is consistent with the quantum prediction and excludes models that preserve both AOE and generalized causality applied to the instrumental Wigner's causal structure. Although the present implementation does not close the detection loophole, it establishes the experimental accessibility of stronger forms of nonclassicality in causal structures beyond Bell.

Beyond the Wigner's friend framework, our results contribute to the broader program of relaxing the assumptions underlying nonclassicality tests in networks. In a general, theory-independent manner, they show that quantum predictions violate assumptions weaker than those required by classical causal models, even in scenarios without spacelike separation. In this sense, the assumptions about reality challenged by quantum correlations are narrowed down (see Ref. [74] for a parallel approach relaxing standard Bell's theorem assumptions). More generally, this work contributes to the broader context of causal networks featuring hybrid OPT-classical assumptions, such as those involving independent sources [35, 75–77] or broadcasting scenarios [68, 78, 79]. Beyond their foundational implications, our results also point towards new applications for quantum technologies that harness violations of CF in general causal structures, particularly in light of the role of LF-violating correlations in theory-independent randomness certification [55].

From a foundational perspective, the violation of CF can only be accommodated by abandoning at least one of the assumptions entering its definition: AOE or one of the two generalized causal constraints associated with the instrumental Wigner's network. Interpretations that deny a single observer-independent account of all observed events, including relational [80], many-worlds [81], QBist [82], and other broadly neo-Copenhagen approaches, would explain the violation by rejecting AOE [83]. By contrast, approaches that retain AOE must be structurally radical, in the sense that they reject the assumed (generalized or classical) causal structure. For instance, de Broglie-Bohm pilot-wave theory [62, 84] allows a direct causal influence from Alice's choice to Bob's outcome. Other approaches reject the mathematical constraint that we have called free choice, Eq. (3), allowing correlations between Alice's choice and Charlie's observed result, either due to retrocausality [85] or superdeterminism [86]. Finally, spontaneous-collapse models [87] may offer a different route if their dynamics prevent the reversible unitary control required by the

protocol, although whether they do so depends on the specific collapse model, its parameters, and the physical implementation of the friend [17].

ACKNOWLEDGMENTS

The authors thank Joshua Laux for useful discussions and comments. MH thanks Ryan Savage for useful discussions concerning the generalized Bayesian networks framework of Ref. [8]. MH acknowledges use of Claude [Sonnet 5], Large Language Model, (Anthropic, 2026, https://www.anthropic.com) for helping with algebra, and suggesting the solution in Eq. (H21) which was verified by the author. This work was supported by Australian Research Council (ARC) Discovery Early Career Researcher Award (DE250100762) (E.P.) and ARC Discovery Project DP25010216 (H.M.W, N.T, E.C). A.D. acknowledges support by the Australian Government Research Training Program (RTP). The authors acknowledge support from FARE Ricerca in Italia QU-DICE Grant n. R20TRHTSPA and by the ERC Advanced Grant QU-BOSS (QUantum advantage via nonlinear BOSon Sampling, grant agreement no. 884676).

[1] J. Pearl, *Causality* (Cambridge university press, 2009).
[2] J. S. Bell, On the Einstein-Podolsky-Rosen paradox, Physics Physique Fizika **1**, 195 (1964).
[3] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Bell nonlocality, Rev. Mod. Phys. **86**, 419 (2014), arXiv:1303.2849 [quant-ph].
[4] C. J. Wood and R. W. Spekkens, The lesson of causal discovery algorithms for quantum correlations: causal explanations of Bell-inequality violations require fine-tuning, New J. Phys. **17**, 033002 (2015), arXiv:1208.4119 [quant-ph].
[5] H. M. Wiseman and E. G. Cavalcanti, Causarum Investigatio and the Two Bell's Theorems of John Bell, in *Quantum [Un]Speakables II: 50 Years of Bell's Theorem: A celebration of John Bell, his theorem and its consequences* (2015) arXiv:1503.06413 [quant-ph].
[6] R. Chaves, R. Kueng, J. B. Brask, and D. Gross, A unifying framework for relaxations of the causal assumptions in Bell's theorem, Phys. Rev. Lett. **114**, 140403 (2015), arXiv:1411.4648 [quant-ph].
[7] E. G. Cavalcanti and R. Lal, On modifications of Reichenbach's principle of common cause in light of Bell's theorem, J. Phys. A **47**, 424018 (2014), arXiv:1311.6852 [quant-ph].
[8] M. F. Pusey, R. Lal, and J. Henson, Theory-independent limits on correlations from generalized Bayesian networks, New J. Phys. **16**, 113043 (2014), arXiv:1405.2572 [quant-ph].
[9] F. Costa and S. Shrapnel, Quantum causal modelling, New J. Phys. **18**, 063032 (2016), arXiv:1512.07106 [quant-ph].
[10] J.-M. A. Allen, J. Barrett, D. C. Horsman, C. M. Lee, and R. W. Spekkens, Quantum Common Causes and Quantum Causal Models, Phys. Rev. X **7**, 031021 (2017), arXiv:1609.09487 [quant-ph].
[11] E. Wolfe, A. Pozas-Kerstjens, M. Grinberg, D. Rosset, A. Acín, and M. Navascues, Quantum Inflation: A General Approach to Quantum Causal Compatibility, Phys. Rev. X **11**, 021043 (2021), arXiv:1909.10519 [quant-ph].
[12] J. Barrett, R. Lorenz, and O. Oreshkov, Quantum Causal Models (2019), arXiv:1906.10726 [quant-ph].
[13] K.-W. Bong, A. Utreras-Alarcón, F. Ghafari, Y.-C. Liang, N. Tischler, E. G. Cavalcanti, G. J. Pryde, and H. M. Wiseman, A strong no-go theorem on the Wigner's friend paradox, Nature Phys. **16**, 1199 (2020), arXiv:1907.05607 [quant-ph].
[14] E. G. Cavalcanti and H. M. Wiseman, Implications of Local Friendliness violation for quantum causality, Entropy **23**, 925 (2021), arXiv:2106.04065 [quant-ph].
[15] Y. Yīng, M. M. Ansanelli, A. Di Biagio, E. Wolfe, D. Schmid, and E. G. Cavalcanti, Relating Wigner's Friend Scenarios to Nonclassical Causal Compatibility, Monogamy Relations, and Fine Tuning, Quantum **8**, 1485 (2024), arXiv:2309.12987 [quant-ph].
[16] M. Haddara and E. G. Cavalcanti, Local friendliness polytopes in multipartite scenarios, Phys. Rev. A **111**, 012206 (2025), arXiv:2407.20346 [quant-ph].
[17] H. M. Wiseman, E. G. Cavalcanti, and E. G. Rieffel, A "thoughtful" Local Friendliness no-go theorem: a prospective experiment with new assumptions to suit, Quantum **7**, 1112 (2023), arXiv:2209.08491 [quant-ph].
[18] D. Schmid, Y. Yīng, and M. Leifer, A review and analysis of six extended Wigner's friend arguments (2023), arXiv:2308.16220 [quant-ph].
[19] A. Balke and J. Pearl, Bounds on treatment effects from studies with imperfect compliance, Journal of the American Statistical Association **92**, 1171 (1997), https://doi.org/10.1080/01621459.1997.10474074.
[20] J. Pearl, On the testability of causal models with latent and instrumental variables, in *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence* (1995) pp. 435–443, arXiv:1302.4976 [cs.AI].
[21] R. Chaves, G. Carvacho, I. Agresti, V. Di Giulio, L. Aolita, S. Giacomini, and F. Sciarrino, Quantum violation of an instrumental test, Nature Phys. **14**, 291 (2018), arXiv:1808.09894 [quant-ph].
[22] T. Van Himbeeck, J. B. Brask, S. Pironio, R. Ramanathan, A. B. Sainz, and E. Wolfe, Quantum violations in the Instrumental scenario and their relations to the Bell scenario, Quantum **3**, 186 (2019), arXiv:1804.04119 [quant-ph].
[23] I. Agresti, D. Poderini, B. Polacchi, N. Miklin, M. Gachechiladze, A. Suprano, E. Polino, G. Milani, G. Carvacho, R. Chaves, and F. Sciarrino, Experimental test of quantum causal influences, Science Advances **8**, eabm1515 (2021), arXiv:2108.08926 [quant-ph].
[24] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Probabilistic theories with purification, Phys. Rev. A **81**, 062348 (2010), arXiv:0908.1583 [quant-ph].
[25] H. Price and R. Corry, *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited* (Oxford University Press, 2007).

[26] C. Rovelli, Back to reichenbach: Back to reichenbach, Journal for General Philosophy of Science / Zeitschrift für Allgemeine Wissenschaftstheorie **56**, 97 (2024).

[27] A. Tavakoli, A. Pozas-Kerstjens, M.-X. Luo, and M.-O. Renou, Bell nonlocality in networks, Rept. Prog. Phys. **85**, 056001 (2022), arXiv:2104.10700 [quant-ph].

[28] C. Branciard, D. Rosset, N. Gisin, and S. Pironio, Bilocal versus nonbilocal correlations in entanglement-swapping experiments, Phys. Rev. A **85**, 032119 (2012), arXiv:1112.4502 [quant-ph].

[29] T. Fritz, Beyond Bell's theorem: Correlation scenarios, New Journal of Physics **14**, 103001 (2012).

[30] M.-O. Renou, E. Bäumer, S. Boreiri, N. Brunner, N. Gisin, and S. Beigi, Genuine Quantum Nonlocality in the Triangle Network, Phys. Rev. Lett. **123**, 140401 (2019), arXiv:1905.04902 [quant-ph].

[31] R. Chaves, G. Moreno, E. Polino, D. Poderini, I. Agresti, A. Suprano, M. R. Barros, G. Carvacho, E. Wolfe, A. Canabarro, R. W. Spekkens, and F. Sciarrino, Causal Networks and Freedom of Choice in Bell's Theorem, PRX Quantum **2**, 040323 (2021), arXiv:2105.05721 [quant-ph].

[32] G. Carvacho, F. Andreoli, L. Santodonato, M. Bentivegna, R. Chaves, and F. Sciarrino, Experimental violation of local causality in a quantum network, Nature Commun. **8**, 14775 (2017), arXiv:1610.03327 [quant-ph].

[33] D. J. Saunders, A. J. Bennet, C. Branciard, and G. J. Pryde, Experimental demonstration of nonbilocal quantum correlations, Sci. Adv. **3**, 1602743 (2017), arXiv:1610.08514 [quant-ph].

[34] D. Poderini, I. Agresti, G. Marchese, E. Polino, T. Giordani, A. Suprano, M. Valeri, G. Milani, N. Spagnolo, G. Carvacho, R. Chaves, and F. Sciarrino, Experimental violation of n-locality in a star quantum network, Nature Communications **11**, 1 (2020).

[35] N.-N. Wang, A. Pozas-Kerstjens, C. Zhang, B.-H. Liu, Y.-F. Huang, C.-F. Li, G.-C. Guo, N. Gisin, and A. Tavakoli, Certification of non-classicality in all links of a photonic star network without assuming quantum mechanics, Nature Commun. **14**, 2153 (2023), arXiv:2212.09765 [quant-ph].

[36] E. Polino, D. Poderini, G. Rodari, I. Agresti, A. Suprano, G. Carvacho, E. Wolfe, A. Canabarro, G. Moreno, G. Milani, R. W. Spekkens, R. Chaves, and F. Sciarrino, Experimental nonclassicality in a causal network without assuming freedom of choice, Nature Commun. **14**, 909 (2023), arXiv:2210.07263 [quant-ph].

[37] X.-M. Gu, L. Huang, A. Pozas-Kerstjens, Y.-F. Jiang, D. Wu, B. Bai, Q.-C. Sun, M.-C. Chen, J. Zhang, S. Yu, Q. Zhang, C.-Y. Lu, and J.-W. Pan, Experimental Full Network Nonlocality with Independent Sources and Strict Locality Constraints, Phys. Rev. Lett. **130**, 190201 (2023), arXiv:2302.02472 [quant-ph].

[38] N.-N. Wang, C. Zhang, H. Cao, K. Xu, B.-H. Liu, Y.-F. Huang, C.-F. Li, G.-C. Guo, N. Gisin, T. Kriváchy, *et al.*, Experimental Genuine Quantum Nonlocality in the Triangle Network, Phys. Rev. Lett. **136**, 180202 (2026), arXiv:2401.15428 [quant-ph].

[39] A. Fine, Insolubility of the quantum measurement problem, Phys. Rev. D **2**, 2783 (1970).

[40] J. Bell, Against 'measurement', Physics world **3**, 33 (1990).

[41] E. P. Wigner, Review of the quantum-mechanical measurement problem, Science, Computers, and the Information Onslaught , 63 (1984).

[42] T. Maudlin, Three measurement problems, Topoi **14**, 7 (1995).

[43] J. Bub and I. Pitowsky, Two dogmas about quantum mechanics (2008), arXiv:0712.4258 [quant-ph].

[44] E. G. Cavalcanti, Bell's theorem and the measurement problem: reducing two mysteries to one? (2016), arXiv:1602.07404 [quant-ph].

[45] Č. Brukner, On the quantum measurement problem, in *Quantum [Un]Speakables II: Half a Century of Bell's Theorem*, edited by R. Bertlmann and A. Zeilinger (Springer International Publishing, Cham, 2017) pp. 95–117, arXiv:1507.05255 [quant-ph].

[46] N. Gisin, Collapse. What else? (2017), arXiv:1701.08300 [quant-ph].

[47] N. Ormrod, V. Vilasini, and J. Barrett, Which theories have a measurement problem?, New J. Phys. **28**, 024506 (2026), arXiv:2303.03353 [quant-ph].

[48] E. P. Wigner, Remarks on the mind-body question, in *Philosophical Reflections and Syntheses*, edited by J. Mehra (Springer Berlin Heidelberg, Berlin, Heidelberg, 1995) pp. 247–260.

[49] E. P. Wigner, The problem of measurement, in *Philosophical Reflections and Syntheses*, edited by J. Mehra (Springer Berlin Heidelberg, Berlin, Heidelberg, 1995) pp. 163–180.

[50] D. Deutsch, Quantum theory as a universal physical theory, Int. J. Theor. Phys. **24**, 1 (1985).

[51] D. Frauchiger and R. Renner, Quantum theory cannot consistently describe the use of itself, Nature Commun. **9**, 3711 (2018), arXiv:1604.07422 [quant-ph].

[52] C. Brukner, A no-go theorem for observer-independent facts, Entropy **20**, 350 (2018), arXiv:1804.00749 [quant-ph].

[53] M. Proietti, A. Pickston, F. Graffitti, P. Barrow, D. Kundys, C. Branciard, M. Ringbauer, and A. Fedrizzi, Experimental test of local observer-independence, Science Advances **5**, eaaw9832 (2019), arXiv:1902.05080 [quant-ph].

[54] Č. Brukner, Wigner's friend and relational objectivity, Nature Reviews Physics **4**, 628 (2022).

[55] E. Woodhead, *Imperfections and self testing in prepare-and-measure quantum key distribution*, Ph.D. thesis, Ph. D. thesis, Université libre de Bruxelles (2014).

[56] J. D. Angrist, G. W. Imbens, and D. B. Rubin, Identification of causal effects using instrumental variables, Journal of the American Statistical Association **91**, 444 (1996).

[57] M. Mogstad, A. Santos, and A. Torgovitsky, Using instrumental variables for inference about policy relevant treatment parameters, Econometrica **86**, 1589 (2018).

[58] B. Bonet, Instrumentality tests revisited, in *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, UAI'01 (Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2001) p. 48–55, arXiv:1301.2258 [cs.AI].

[59] A. B. Sainz, M. J. Hoban, P. Skrzypczyk, and L. Aolita, Bipartite post-quantum steering in generalised scenarios, Phys. Rev. Lett. **125**, 050404 (2020), arXiv:1907.03705 [quant-ph].

[60] P. J. Cavalcanti, J. H. Selby, J. Sikora, T. D. Galley, and A. B. Sainz, Post-quantum steering is a stronger-than-quantum resource for information processing, npj Quantum Inf. **8**, 76 (2022), arXiv:2102.06581 [quant-ph].

[61] M. Gachechiladze, N. Miklin, and R. Chaves, Quantifying causal influences in the presence of a quantum common cause, Phys. Rev. Lett. **125**, 230401 (2020), arXiv:2007.01221 [quant-ph].

[62] D. Bohm and B. J. Hiley, *The Undivided Universe: An Ontological Interpretation of Quantum Theory* (Routledge, 1995).

[63] E. G. Cavalcanti, The view from a Wigner bubble, Found. Phys. **51**, 39 (2021), arXiv:2008.05100 [quant-ph].

[64] E. Adlam, What Does '(Non)-absoluteness of Observed Events' Mean?, Found. Phys. **54**, 13 (2024), arXiv:2309.03171 [quant-ph].

[65] Allowing a stochastic classical reporting channel does not allow correlations that are not reproducible by conditional determinism. Indeed, any classical channel $T(a|c)$ can be represented as a deterministic function $a = f(c, r)$ of its input and an auxiliary classical random variable R. Absorbing this randomness into an enlarged latent description therefore makes Alice's response at $x = 0$ deterministic. Defining the effective classical record $C' = f(C, R)$ recovers the tracking relation $P(a|x = 0, c') = \delta_{a,c'}$, although C' need not coincide with Charlie's original record C.

[66] A. Cabello, Specker's fundamental principle of quantum mechanics (2012), arXiv:1212.1756 [quant-ph].

[67] A. Cabello, Simple explanation of the quantum violation of a fundamental inequality, Phys. Rev. Lett. **110**, 060402 (2013), arXiv:1210.2988 [quant-ph].

[68] G. Rodari, D. Poderini, E. Polino, A. Suprano, F. Sciarrino, and R. Chaves, Characterizing Hybrid Causal Structures with the Exclusivity Graph Approach (2023), arXiv:2401.00063 [quant-ph].

[69] Note that the notion of Causal Friendliness introduced here is distinct from the "Causal Time-Symmetric Friendliness" considered in Ref. [88], which combines AOE with other assumptions named Axiological Time Symmetry, No Retrocausality, and Screening via Pseudo Events in a sequential prepare-and-measure configuration.

[70] M. Haddara, H. M. Wiseman, and E. G. Cavalcanti, Properties and Applications of Partially Deterministic Polytopes (2025), arXiv:2510.25127 [quant-ph].

[71] A. Acín, S. Pironio, and M. Navascués, A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations, New J. Phys. **10**, 073013 (2008), arXiv:0803.4290 [quant-ph].

[72] A. Fedrizzi, T. Herbst, A. Poppe, T. Jennewein, and A. Zeilinger, A wavelength-tunable fiber-coupled source of narrowband entangled photons, Opt. Express **15**, 15377 (2007), arXiv:0706.2877 [quant-ph].

[73] G. Moreno, R. Nery, C. Duarte, and R. Chaves, Events in quantum mechanics are maximally non-absolute, Quantum **6**, 785 (2022), arXiv:2112.11223 [quant-ph].

[74] C. Vieira, R. Ramanathan, and A. Cabello, Test of the physical significance of Bell non-locality, Nature Commun. **16**, 4390 (2025), arXiv:2402.00801 [quant-ph].

[75] A. Pozas-Kerstjens, N. Gisin, and A. Tavakoli, Full Network Nonlocality, Phys. Rev. Lett. **128**, 010403 (2022), arXiv:2105.09325 [quant-ph].

[76] M. C. Alañón, E.-C. Boghiu, P. Abiuso, and E. Wolfe, Escaping the Shadow of Bell's Theorem in Network Nonlocality, Quantum **10**, 2213 (2026), arXiv:2406.15587 [quant-ph].

[77] E. Polino, D. Poderini, G. Minati, G. Rodari, R. Chaves, and F. Sciarrino, Network Nonlocality with Separable Measurements (2026), arXiv:2604.11910 [quant-ph].

[78] J. Bowles, F. Hirsch, and D. Cavalcanti, Single-copy activation of Bell nonlocality via broadcasting of quantum states, Quantum **5**, 499 (2021), arXiv:2007.16034 [quant-ph].

[79] L. Villegas-Aguilar, E. Polino, F. Ghafari, M. T. Quintino, K. T. Laverick, I. R. Berkman, S. Rogge, L. K. Shalm, N. Tischler, E. G. Cavalcanti, S. Slussarenko, and G. J. Pryde, Nonlocality activation in a photonic quantum network, Nature Commun. **15**, 3112 (2024), arXiv:2309.06501 [quant-ph].

[80] C. Rovelli, Relational quantum mechanics, International Journal of Theoretical Physics **35**, 1637 (1996).

[81] L. Vaidman, The many-worlds interpretation of quantum mechanics: Current status and relation to other interpretations (2024).

[82] C. A. Fuchs and B. C. Stacey, QBism: Quantum theory as a hero's handbook, Proc. Int. Sch. Phys. Fermi **197**, 133 (2019).

[83] D. Schmid, Y. Yīng, and M. Leifer, Copenhagenish interpretations of quantum mechanics (2025), arXiv:2506.00112 [quant-ph].

[84] D. Duerr and S. Teufel, *Bohmian mechanics. The physics and mathematics of quantum theory* (Springer, 2009).

[85] K. B. Wharton and N. Argaman, Colloquium : Bell's theorem and locally mediated reformulations of quantum mechanics, Rev. Mod. Phys. **92**, 021002 (2020), arXiv:1906.04313 [quant-ph].

[86] S. Hossenfelder and T. N. Palmer, Rethinking Superdeterminism, Front. in Phys. **8**, 139 (2020), arXiv:1912.06462 [quant-ph].

[87] A. Bassi, K. Lochan, S. Satin, T. P. Singh, and H. Ulbricht, Models of Wave-function Collapse, Underlying Theories, and Experimental Tests, Rev. Mod. Phys. **85**, 471 (2013), arXiv:1204.4325 [quant-ph].

[88] S. Mukherjee and J. R. Hance, Limits of absoluteness of observed events in timelike scenarios: A no-go theorem, Phys. Rev. Res. **8**, 023066 (2026), arXiv:2510.26562 [quant-ph].

[89] A. Cabello, S. Severini, and A. Winter, Graph-Theoretic Approach to Quantum Correlations, Phys. Rev. Lett. **112**, 040401 (2014), arXiv:1401.7081 [quant-ph].

[90] A. Acín, T. Fritz, A. Leverrier, and A. B. Sainz, A Combinatorial Approach to Nonlocality and Contextuality, Commun. Math. Phys. **334**, 533 (2015), arXiv:1212.4084 [quant-ph].

[91] D. E. Knuth, The sandwich theorem (1993), arXiv:math/9312214 [math.CO].

[92] M. Grötschel, L. Lovász, and A. Schrijver, Relaxations of vertex packing, Journal of Combinatorial Theory, Series B **40**, 330 (1986).

[93] A. Cabello, L. E. Danielsen, A. J. Lopez-Tarrida, and J. R. Portillo, Basic exclusivity graphs in quantum correlations, Phys. Rev. A **88**, 032104 (2013), arXiv:1211.5825 [quant-ph].

[94] D. Poderini, R. Chaves, I. Agresti, G. Carvacho, and F. Sciarrino, Exclusivity graph approach to instrumental inequalities, in *Uncertainty in Artificial Intelligence* (PMLR, 2020) pp. 1274–1283, arXiv:1909.09120 [quant-ph].

[95] The Sage Development Team, *SageMath: Combinatorial and Discrete Geometry*, SageMath (2025), polyhedra and

combinatorial geometry module.

[96] A. Cabello, A. Feito, and A. Lamas-Linares, Bell's inequalities with realistic noise for polarization-entangled photons, Phys. Rev. A **72**, 052112 (2005), arXiv:quant-ph/0509027.

[97] G. Cañas, J. F. Barra, E. S. Gómez, G. Lima, F. Sciarrino, and A. Cabello, Detection efficiency for loophole-free Bell tests with entangled states affected by colored noise, Phys. Rev. A **87**, 012113 (2013), arXiv:1206.2290 [quant-ph].

[98] M. J. W. Hall, Complementary contributions of indeterminism and signalling to quantum correlations, Phys. Rev. A **82**, 062117 (2010), arXiv:1006.3680 [quant-ph].

[99] G. Chiribella, L. Giannelli, and C. M. Scandolo, Bell Nonlocality in Classical Systems Coexisting with Other System Types, Phys. Rev. Lett. **132**, 190201 (2024), arXiv:2301.10885 [quant-ph].

APPENDIX/SUPPLEMENTARY MATERIAL

Appendix A: Operational Probabilistic Theories for generalized causal structures

To model nonclassical causal relations, we adopt the generalized Bayesian-network framework of Henson *et al.* [8], formulated in the language of OPTs [24]. In an OPT, laboratory procedures are represented by tests that can be composed into circuits, with the theory assigning probabilities to the outcomes of any such composition. The corresponding notion of causality is weaker than the classical one: it only requires that the compositionality of the underlying OPT respects the causal structure in an appropriate way. This implies, e.g., that the statistics of a subset of variables does not depend on choices that are not among their causes, *i.e.*, there is no-signaling from the causal future to the causal past. It also implies that correlations between variables with no directed causal path between them must be explained in terms of a common cause – what in [7] was called the Principle of Common Cause – but does not imply *Reichenbach's* Principle of Common Cause, which further requires a classical explanation whereby conditioning on the common cause eliminates the correlation. This framework naturally accommodates quantum theory, as well as more general probabilistic models.

A generalized causal structure is specified by a directed acyclic graph whose nodes are separated into observed and unobserved classes. The observed nodes encode classical random variables, and arrows coming out of observed nodes (observed edges) are associated with a classical copy of that node's variable. Unobserved nodes represent preparations or transformations of systems in the OPT: each arrow coming out of an unobserved node (unobserved edges) is associated with a distinct system S_{X^i} in the theory, and the particular operation realized at an unobserved node may depend on its classical inputs.

More formally, for each node $X^{(i)}$, and for each assignment $\mathrm{opa}\, x^{(i)}$ of its observed parents, one introduces a local *test* $\mathcal{T}_{x^{(i)}}(\mathrm{opa}\, x^{(i)})^{\mathrm{outU}X^{(i)}}_{\mathrm{incU}X^{(i)}}$, which transforms the composite system $\mathrm{incU}X^{(i)}$ entering the node from its incoming unobserved edges (if any) into the composite system $\mathrm{outU}X^{(i)}$ leaving the node through outgoing unobserved edges (if any). The subscript $x^{(i)}$ represents the outcome of the test. When $X^{(i)}$ is an observed node, the outcomes of the associated test $\mathcal{T}_{x^{(i)}}(\mathrm{opa}\, x^{(i)})_{\mathrm{incU}X^{(i)}}$ (which has no outgoing systems) coincide with the possible values $x^{(i)}$ of the variable $X^{(i)}$; when $X^{(i)}$ is an unobserved node, the test $\mathcal{T}(\mathrm{opa}\, x^{(i)})^{\mathrm{outU}X^{(i)}}_{\mathrm{incU}X^{(i)}}$ has a single trivial outcome (which can be omitted from the subscript). The observed joint distribution is obtained by composing all such local tests in the pattern dictated by the graph.

A probability distribution over the observed variables is *compatible* with the causal structure if and only if there exists an assignment of tests to the nodes satisfying the rules above such that it satisfies the generalized Markov condition [8]:

$$P\big(x^{(1)}, x^{(2)}, \ldots, x^{(n)}\big) = \prod_{i=1}^{m} \mathcal{T}_{x^{(i)}}(\mathrm{opa}\, x^{(i)})^{\mathrm{outU}X^{(i)}}_{\mathrm{incU}X^{(i)}}. \tag{A1}$$

In the special case when all variables in the structure are observed, all tests reduce to conditional probabilities $P(x^{(i)}|\mathrm{opa}\, x^{(i)})$. In this way, the classical Markov condition is recovered, while the generalized graphical language can be used to describe nonclassical causal models.

For more details of the OPT theory-independent causal framework, we refer to Ref. [8].

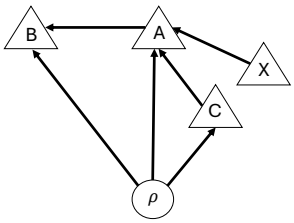


FIG. 4: A generalized DAG, in the graphical notation of [8], corresponding to the instrumental-based causal structure where the CF no-go result is exhibited. The triangles represent observed nodes and have random variables assigned to them, while the circle represents a latent node which is associated with some general OPT systems but produces no outcomes.

Appendix B: Necessary and sufficient constraints from the instrumental Wigner's friend causal structure

In this section, we will show that Pearl's inequalities in Eq. (4) and the freedom of choice constraints in Eq. (3) of the main text capture all the relevant causal constraints on $P(a, b, c|x)$ in the CF-instrumental scenario, and hence when paired with the tracking assumption in Eq. (2), they also capture all the relevant constraints for Causal Friendliness in the scenario, leading to the full characterization of the CF polytope presented in the main text.

In the case of the CF-instrumental scenario where the causal structure is as portrayed in Fig. 4 with the same notation used in [8], the generalized causal Markov condition of Eq. (A1) gives

$$\begin{aligned} P(a,b,c,x) &= \mathcal{T}_b(a)^I_{S_B}\mathcal{T}_a(x,c)^I_{S_A}\mathcal{T}_c()^I_{S_C}\mathcal{T}_x()^I_I\mathcal{T}_\rho()^{S_AS_BS_C}_I \\ &= \mathcal{T}_b(a)^I_{S_B}\mathcal{T}_a(x,c)^I_{S_A}\mathcal{T}_c()^I_{S_C}\mathcal{T}_\rho()^{S_AS_BS_C}_I P(x), \end{aligned} \tag{B1}$$

where, on the second line, the fact that a test from a trivial system to another one is associated with a probability is used. Here $\mathcal{T}_\rho()^{S_AS_BS_C}_I$ is the preparation-test associated with the unobserved source, $\mathcal{T}_c()^I_{S_C}$ is the outcome-c

event at C, $\mathcal{T}_a(x,c)^I_{S_A}$ is the outcome-a event at A, chosen as a function of the observed parents x, c, and $\mathcal{T}_b(a)^I_{S_B}$ is the outcome-b event at B, chosen as a function of the observed parent a. The systems S_A, S_B, S_C refer to the ones that connect the preparation test to the tests at the nodes A, B, C respectively.

We can now compute the probability $P(c,x)$ of the free choice x and the classical outcome c by summing over the outcomes a and b:

$$\begin{aligned} P(c,x) &= \sum_{a,b} P(a,b,c,x) \\ &= P(x) \sum_{a,b} \mathcal{T}_b(a)^I_{S_B} \mathcal{T}_a(x,c)^I_{S_A} \mathcal{T}_c()^I_{S_C} \mathcal{T}_\rho()^{S_A S_B S_C}_I . \end{aligned} \tag{B2}$$

The sums over a and b are local coarse-grainings of complete observation-tests. Therefore, by OPT causality:

$$\sum_a \mathcal{T}_a(x,c)^I_{S_A} = \top^I_{S_A}, \qquad \sum_b \mathcal{T}_b(a)^I_{S_B} = \top^I_{S_B} , \tag{B3}$$

where, following the notation in [8], $\top^I_{S_A}$ and $\top^I_{S_B}$ are the unique deterministic effects on S_A and S_B. It follows that

$$P(c,x) = P(x) \top^I_{S_B} \top^I_{S_A} \mathcal{T}_c()^I_{S_C} \mathcal{T}_\rho()^{S_A S_B S_C}_I = P(x)P(c), \tag{B4}$$

giving the No-Superdeterminism/Freedom of Choice condition of Eq.(3).

Using $P(a,b,c,x) = P(a,b|c,x)P(c,x)$, the remaining relevant constraints are found from $P(a,b|x,c)$. Note first that (assuming $P(x,c) > 0$)

$$P(a,b \mid x,c) = \frac{\mathcal{T}_b(a)^I_{S_B} \mathcal{T}_a(x,c)^I_{S_A} \mathcal{T}_c()^I_{S_C} \mathcal{T}_\rho()^{S_A S_B S_C}_I}{P(c)}. \tag{B5}$$

We will next show that any family of distributions $P(a,b|x,c)$ is compatible with the decomposition in Eq. (B5) if and only if they obey Pearl's inequalities as stated in Eq. (4). This is in fact, an immediate corollary of the results in Refs. [8, 22] where constraints on the *ordinary* instrumental scenario (i.e., the one which corresponds to the causal structure of Fig. 4 without the C node) were investigated. The same techniques can be adapted to our case.

The trick is to notice (following Refs. [8, 22]) the close similarity of the constraints in Eq. (B5) to those arising in a *canonical Local Friendliness scenario* (in the terminology of Ref. [16]) where Alice and Bob are spacelike separated and the arrow from A to B is absent. The causal structure corresponding to this scenario is illustrated in Fig. 5. The generalized causal Markov condition of Eq. (A1) implies that a distribution $Q(a,b,c,x,y)$ compatible with the causal structure in Fig. 5 decomposes as

$$\begin{aligned} Q(a,b,c,x,y) = \mathcal{T}'_b(y)^I_{S'_B} \mathcal{T}'_a(x,c)^I_{S'_A} \mathcal{T}'_c()^I_{S'_C} \mathcal{T}'_\rho()^{S'_A S'_B S'_C}_I \\ \times Q(x)Q(y). \end{aligned} \tag{B6}$$

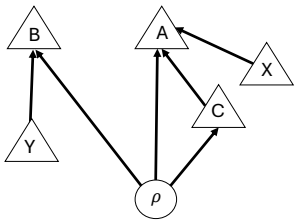


FIG. 5: The generalized DAG corresponding to a canonical Local Friendliness scenario. In this picture, Alice and Bob perform test at spacelike separation, which corresponds to ruling out causal influence from the events on Bob's side to Alice's, and vice versa.

The analog argument shown for the case of the instrumental scenario implies that $Q(x,y,c) = Q(x)Q(y)Q(c)$ and so (with $Q(x,y,c) > 0$) that

$$Q(a,b|x,y,c) = \frac{\mathcal{T}'_b(y)^I_{S'_B} \mathcal{T}'_a(x,c)^I_{S'_A} \mathcal{T}'_c()^I_{S'_C} \mathcal{T}'_\rho()^{S'_A S'_B S'_C}_I}{Q(c)}. \tag{B7}$$

The most general (i.e., complete) constraints on $Q(a,b|x,y,c)$ are well understood in these type of scenarios [13, 15, 16, 70]; they essentially demand that the $Q(a,b|x,y,c)$ satisfy parameter independence $\sum_a Q(a,b|x,y,c) = Q(b|y,c)$ and $\sum_b Q(a,b|x,y,c) = Q(a|x,c)$ for all x,y,c which is to say that the collections of distributions $Q(a,b|x,y,c)$ are members of the no-signaling polytope for every c.

The observation that connects the above discussion back to the instrumental scenario is that since the tests and systems are allowed to be arbitrary, a comparison between the two scenarios *within the same theory* allows one to equate the systems S_A, S_B, S_C in the instrumental case with the primed systems in the canonical scenario, and permits the applications of the same tests in both cases, so that $\mathcal{T}'_c()^I_{S'_C} = \mathcal{T}_c()^I_{S_C}$ and so on. Therefore, without loss of generality, the distributions $Q(a,b|x,y,c)$ are compatible with generalized causal Markov in the scenario depicted in Fig. 5 if and only if

$$Q(a,b|x,y,c) = \frac{\mathcal{T}'_b(y)^I_{S_B} \mathcal{T}_a(x,c)^I_{S_A} \mathcal{T}_c()^I_{S_C} \mathcal{T}_\rho()^{S_A S_B S_C}_I}{P(c)}. \tag{B8}$$

Now it is evident that if the cardinality of the set Y is at least as great as that of A, it is possible to reproduce exactly the type of constraints on $P(a,b|x,c)$ as expressed in Eq. (B5) by considering the constraints on distributions $Q(a,b|x,y,c)$ as in Eq. (B8) and setting $\mathcal{T}'_b(y = y_a) = \mathcal{T}_b(a)$ for some $y_a \in Y$ (since the labels are arbitrary, without loss of generality $y_a = a$). The claim *is not* that such a "choice" of y can be realized in the canonical scenario given the output a (since communication is excluded), but rather that the constraints

in Eq. (B5) are inherently built in on the subsets of the distributions of the behaviours $Q(a,b|x,y,c)$, namely the ones of the form $Q(a,b|x,y=a,c)$, so that identification

$$\begin{aligned} P(a,b|x,c) =& Q(a,b|x,y=a,c) \\ & \forall\ x,\ a,\ b,\ c \end{aligned} \tag{B9}$$

is possible. Eq. (B9) essentially states that the valid instrumental distributions $P(a,b|x,c)$ are exactly those that can be recovered from behaviors $Q(a,b|x,y,c)$ by a mapping that only includes the probabilities where $y=a$. This map can also be directly seen by using the diagrammatic representations of the generalized causal structures [8]: the diagram of the causal structure in Fig. 4 matches exactly with the one corresponding to the causal structure in Fig. 5 when the condition $y=a$ is set. In Ref. [22], it was pointed out that this map essentially amounts to projecting the no-signaling polytope (where the distributions $Q(a,b|x,y,c)$ live) onto a subspace, yielding another polytope.

As we show in Appendix C, the (non-trivial) facets of the projected polytope, which therefore capture the complete constraints for the instrumental distributions $P(a,b|x,c)$ are exactly the Pearl's inequalities (the case for the variables cardinalities $|X|=3, |A|=|B|=2$, was already shown in Ref. [22]), which completes the proof.

As an immediate corollary, the CF-polytope is fully characterized by the causal constraints which are obeyed by all OPTs in the instrumental-based scenario, which are expressed in terms of Pearl's inequalities in Eq. (4) and Freedom of Choice in Eq. (3), and the tracking constraint in Eq. (2) which is enforced by the specific reading of the measurement labelled by X. Let us also mention that the tracking constraint can also be directly baked into the distributions $Q(a,b|x,y,c)$ by demanding that they belong to the partially deterministic polytope [16, 70] corresponding to the canonical Local Friendliness scenario of Fig. 5, and projecting those to the instrumental scenario via Eq. (B9). By either approach, the complete characterization of the CF instrumental polytope presented in Sec. III is obtained.

Appendix C: Uniqueness of Pearl's inequalities for the OPT instrumental scenario

Pearl's inequalities (conditioned on c) in Eq. (4) of the main text can be compactly expressed as

$$\max_a \sum_b \max_x P(a,b|x,c) \leq 1. \tag{C1}$$

This form of the inequalities recovers the specific expression in Eq. (4) of the main text with the fixed cardinalities, but it can also be evaluated in more general settings as we do in this section.

In appendix B it was established that the distributions $P(a,b,c,x)$ defined over the instrumental-based scenario are those which factor as

$$P(a,b,c,x) = P(a,b|x,c)P(x)P(c) \tag{C2}$$

and where the conditional distributions $P(a,b|x,c)$ are those that can be obtained from a no-signaling behaviour $Q(a,b|x,y,c)$ via the map in Eq. (B9) as $P(a,b|x,c) = Q(a,b|x,y=a,c)$. Here the set Y of Bob's inputs is without loss of generality equal to Alice's outcome set A. In this section, we fully characterize the set of OPT instrumental distributions $P(a,b|x,c)$ compatible with such constraints, demonstrating the following theorem:

Theorem 2. *The OPT instrumental distributions $P(a,b|x,c)$, for any finite cardinality of a,b,x,c, are exactly those satisfying positivity, normalization, and Pearl's instrumental inequalities in the form of Eq. (C1).*

It is sufficient to show that the relation

$$P(a,b|x,c) = Q(a,b|x,y=a,c) \tag{C3}$$

implies that Pearl's inequalities in the form (C1) are satisfied by the instrumental distributions $P(a,b|x,c)$ and conversely, that if Pearl's inequalities hold for $P(a,b|x,c)$ then there exists a no-signaling distribution $Q(a,b|x,y,c)$ such that the relation (C3) holds. In the forthcoming discussion, we refer to Eq. (C3) as "the instrumental constraints".

[$\Longrightarrow$] *Instrumental constraints imply Pearl's inequalities.* Let $Q(a,b|x,y,c)$ be a no-signaling behaviour which recovers the instrumental distributions $P(a,b|x,c)$ via Eq. (C3). Fix a value $y=\alpha$ of Bob's input. Since Q is no-signaling from Alice to Bob, the marginal distribution of Bob's outcome for the input $y=\alpha$ cannot depend on x. Hence, for each b, there exists a number $m^c_{\alpha b}$, independent of x, such that

$$\sum_a Q(a,b|x,y=\alpha,c) = m^c_{\alpha b} \qquad \forall\, x. \tag{C4}$$

In particular, among the terms in the sum over Alice's output a, there is the term with $a=\alpha$. Using the defining relation between P and Q, we have

$$\begin{aligned} m^c_{\alpha b} =& \sum_a Q(a,b|x,y=\alpha,c) \geq Q(\alpha,b|x,y=\alpha,c) \\ =& P(\alpha,b|x,c) \quad \forall\, x,c\,. \end{aligned} \tag{C5}$$

Thus in particular

$$m^c_{\alpha b} \geq \max_x P(\alpha,b|x,c). \tag{C6}$$

Moreover, for each fixed α, c, the numbers $m^c_{\alpha b}$ by definition give a normalized probability distribution over b: $\sum_b m^c_{\alpha b} = \sum_b \sum_a Q(a,b|x,\alpha,c) = 1$. Combining the previous two observations gives

$$\sum_b \max_x P(\alpha,b|x,c) \leq \sum_b m^c_{\alpha b} = 1. \tag{C7}$$

Since this conclusion holds for every value α, it holds in particular for the α which maximizes the sum, so that one obtains Pearl's inequalities:

$$\max_a \sum_b \max_x P(a,b|x,c) \leq 1\ . \tag{C8}$$

This direction was shown by Henson et al. [8].

[$\Longleftarrow$] *Pearl's inequalities imply the instrumental constraints.*
Conversely, suppose that $P(a,b|x,c)$ satisfies Pearl's inequalities so that

$$\sum_b \max_x P(\alpha,b|x,c) \leq 1 \qquad \forall\alpha. \tag{C9}$$

For each value α of Alice's output a, we define a probability distribution $\{m^c_{\alpha b}\}_b$ such that

$$m^c_{\alpha b} \geq \max_x P(\alpha,b|x,c), \qquad \sum_b m^c_{\alpha b} = 1. \tag{C10}$$

Such $m^c_{\alpha b}$ exist precisely because Pearl's inequality holds. As a concrete example, if $\sum_b \max_x P(\alpha,b|x,c) = 1$, the identification $m^c_{\alpha b} = \max_x P(\alpha,b|x,c)$ is sufficient, otherwise if $\sum_b \max_x P(\alpha,b|x,c) = k < 1$, one can define $m^c_{\alpha b'} = \max_x P(\alpha,b'|x,c) + 1 - k$ for some b', and $m^c_{\alpha b} = \max_x P(\alpha,b|x,c)$ for $b \neq b'$, say.

We now want to construct a bipartite no-signaling distribution $Q(a,b|x,y,c)$ satisfying $Q(a,b|x,y=a,c) = P(a,b|x,c)$.
First, with $y=\alpha$, set:

$$Q(\alpha,b|x,y=\alpha,c) = P(\alpha,b|x,c). \tag{C11}$$

We will show that the entries $Q(a,b|x,y=\alpha,c)$ with $a \neq \alpha$ can always be chosen consistently. For fixed x and $y=\alpha$, regard these entries as elements $M^{x\alpha c}_{ab}$ of an unknown non-negative matrix $M^{x\alpha c}$ whose rows are labeled by $a \neq \alpha$ and whose columns are labeled by b. We require its row sums to be

$$R^{x\alpha c}_a = \sum_b Q(a,b|x,y=\alpha,c) := P(a|x,c), \qquad a \neq \alpha, \tag{C12}$$

and its column sums to be

$$C^{x\alpha c}_b = \sum_{a\neq\alpha} Q(a,b|x,y=\alpha,c) := m^c_{\alpha b} - P(\alpha,b|x,c)\ . \tag{C13}$$

The row sums are non-negative because they are probabilities. The column sums are also non-negative because

$$m^c_{\alpha b} \geq \max_x P(\alpha,b|x,c) \geq P(\alpha,b|x,c)\ \forall\alpha,b,x,c. \tag{C14}$$

For these requirements to be consistent, it is necessary that the sum over all the rows equals the sum over all the columns. This is easily verified as the total row sum gives:

$$\sum_{a\neq\alpha} R^{x\alpha c}_a = \sum_{a\neq\alpha} P(a|x,c) = 1 - P(\alpha|x,c)\ , \tag{C15}$$

while

$$\begin{aligned}\sum_b C^{x\alpha c}_b &= \sum_b [m^c_{\alpha b} - P(\alpha,b|x,c)] = \\ &= 1 - \sum_b P(\alpha,b|x,c) = 1 - P(\alpha|x,c)\ .\end{aligned} \tag{C16}$$

This consistency condition is also sufficient for the existence of a completion of the non-negative matrix $M^{x\alpha c}$ of the $Q(a,b|x,y=\alpha,c), a\neq\alpha$ with entries that are compatible with the constraints on the row and column sums in Eqs. (C15)–(C16). Indeed, as a concrete example, one can set

$$M^{x\alpha c}_{ab} := \begin{cases} \dfrac{R^{x\alpha c}_a\, C^{x\alpha c}_b}{1 - P(\alpha \mid x,c)}, & P(\alpha \mid x,c) < 1, \\ 0, & P(\alpha \mid x,c) = 1, \end{cases} \qquad a \neq \alpha\,, \tag{C17}$$

which is always possible. Finally, we observe that every matrix $M^{x\alpha c}$ compatible with the row and column constraints gives a valid extension to a no-signaling distribution $Q(a,b|x,y,c)$. Namely since now by construction

$$\begin{aligned}\sum_a Q(a,b|x,y=\alpha,c) &= C^{x\alpha c}_b + Q(\alpha,b|x,y=\alpha,c) \\ &= [m^c_{\alpha b} - P(\alpha,b|x,c)] + P(\alpha,b|x,c) = m^c_{\alpha b}\ ,\end{aligned} \tag{C18}$$

which is independent of x, so No Signaling (NS) from Alice to Bob holds. For $a=\alpha$ we have

$$\sum_b Q(\alpha,b|x,y=\alpha,c) = \sum_b P(\alpha,b|x,c) = P(\alpha|x,c)\ , \tag{C19}$$

while for $a\neq\alpha$, the prescribed row sums give

$$\sum_b Q(a,b|x,y=\alpha,c) = P(a|x,c)\ . \tag{C20}$$

Thus, for any value of a, also no-signalling from Bob to Alice holds. Therefore Q is a bipartite NS extension of P, and by construction

$$Q(a,b|x,y=a,c) = P(a,b|x,c), \tag{C21}$$

establishing the claim.

Note that we include c among the conditioning variables in the probabilities to account for the presence of the variable C in the instrumental Wigner's friend causal structure shown in Fig. 4. However, exactly the same construction also applies to the standard instrumental scenario, where the node C is absent. Therefore, Pearl's inequalities in the form of Eq. (C1) (without the conditioning on c) give the necessary and sufficient constraints for distributions $P(a,b|x)$ to be compatible with the generalized causal Markov condition in the ordinary instrumental scenario.

Appendix D: Exclusivity principle as an alternative nonclassical principle for generalized causality

In this appendix, we focus on the relation between CF and the exclusivity principle in the instrumental scenario.

The exclusivity graph approach [89, 90], suitably extended, is an alternative technique for characterizing correlations arising from sets of hybrid assumptions in arbitrary scenarios [68]. This approach is particularly convenient when one wants to study or find inequalities satisfied by the models under consideration, without fully characterizing the polytope.

For the instrumental scenario considered in this work, we have dichotomic outcomes for both A and B, i.e., $a, b \in \{-1, +1\}$, and the measurement choice value $x \in \{0, 1, 2\}$. We apply the CF constraints, where the outcome a, conditioned on the value c, is deterministic (more generally, classical) only for the measurement choice $X = 0$. Here, each possible observable event $e = (a, b|x)$ represents a node of the exclusivity graph, and its exclusivity relation with another event $e' = (a', b'|x')$ is represented by an edge and defined as:

$$(a, b|x) \sim (a', b'|x') \iff \begin{cases} a \neq a', \ x = x', \\ \text{or} \\ b \neq b', \ a = a' \ . \end{cases}$$

Notably, the exclusivity principle applied to the instrumental exclusivity graph yields exactly Pearl's inequalities in Eq. (4), thereby fully characterizing the corresponding OPT relations. The complete exclusivity graph for the considered instrumental Wigner's friend scenario is depicted in Fig. 6a. Here, in order to impose different constraints to the events (the AOE and tracking assumption only constrains the events with the choice $X = 0$) there are two kinds of edges: those relative to the constraint of generalized causality in Eq. (4) (blue edges in Fig. 6) which are only constrained by the exclusivity principle, and those connecting exclusive events involved in the AOE relation in Eq. (1), which have deterministic (more generally, classical) outcomes occurring when the choice $X = 0$ (red edges in Fig. 6). The approach in Ref. [68] can handle this kind of graph with hybrid assumptions, retrieving constraints on the observed probabilities from the CF assumptions. From the complete graph, we can select a subgraph (the exclusivity graph of the subset of nodes and their relative exclusivity relations) shown in Fig. 6(b), which represents the graph of the events involved in the genuine facet in Eq. (5). By suitably calculating the graph invariants, we can directly compute the bounds of linear inequalities for the CCM, CF, and OPT models in the instrumental Wigner's friend scenario, without explicitly characterizing the full polytopes.

Given a graph G and weights w on the events and a linear combination of the probabilities $I_{inst}(G, w)$, one can always compute: $\alpha(G, w)$, known as the independence number of a graph, provides a tight bound for classical causal models (CCMs) respecting classical causality; $\theta(G, w)$, called Lovaśz Theta of a graph, provides an upper quantum bound (Q) for the inequality [91, 92] when it is mapped onto an exclusivity graph G and a set of weights w [67, 93, 94]; $\alpha_{EP}(G, w)$ is the bound for OPTs obeying the exclusivity principle, which in standard bipartite scenarios corresponds to NS bounds [89, 90]; $\alpha_{CF}(G, w)$ accounts for the two different types of edges from CF assumptions applied to the instrumental Wigner's friend graph and can be calculated using the approach in Ref. [68].

In general, the relation holds $I(G, w) \overset{\text{CCM}}{\leq} \alpha(G, w) \overset{\text{Q}}{\leq} \theta(G, w) \overset{\text{EP}}{\leq} \alpha_{EP}(G, w)$ The following inequalities also hold $I(G, w) \overset{\text{CCM}}{\leq} \alpha(G, w) \overset{\text{CF}}{\leq} \alpha_{CF}(G, w) \overset{\text{EP}}{\leq} \alpha_{EP}(G, w)$.

Nevertheless, the relation between α_{CF} and $\theta(G, w)$ is not fixed, but depends on the particular inequality considered, given that there is no inclusion relation between CF and quantum correlation sets.

Appendix E: Correlation sets and axes expressions

The expressions on the axes of the 2-D polytopes' projection in Fig. 3 are the following:

$$\begin{aligned} \mathcal{E}_1 &= 2.7 \cdot (\mathcal{E}'_1 + \mathcal{E}'_2) \\ \mathcal{E}_2 &= 2 \cdot (-\mathcal{E}'_1 + \mathcal{E}'_2) \end{aligned} \tag{E1}$$

with:

$$\begin{aligned} \mathcal{E}'_1 = &- \wp(-1, +1|0) + \wp(+1, -1|1) + \wp(-1, +1|1) \\ &+ \wp(+1, +1|2) + \wp(-1, +1|2) \end{aligned} \tag{E2}$$

and

$$\begin{aligned} \mathcal{E}'_2 = &\ 0.2 \cdot \wp(-1, -1|0) + \wp(+1, +1|0) + \wp(-1, -1|2) \\ &+ \wp(+1, -1|2) + 2 \cdot \wp(-1, +1|0) + \wp(+1, +1|1) \\ &+ \wp(-1, -1|1) + \wp(+1, -1|0) \end{aligned} \tag{E3}$$

The choice of the expressions themselves was optimized to showcase the location of the two experimental points; one (X_A) lying entirely within the CF polytope but outside the CCM one, while the other (X_B) lying outside both the CF and the CCM polytopes, as is evident in Fig. 3. The polytopes are computed using the standard Python software SageMath with the Polyhedra library [95].

Appendix F: Quantum strategy violating CF constraints and noise robustness

As mentioned previously in the main text, an optimal quantum strategy to find a violation of Eq. (5) requires

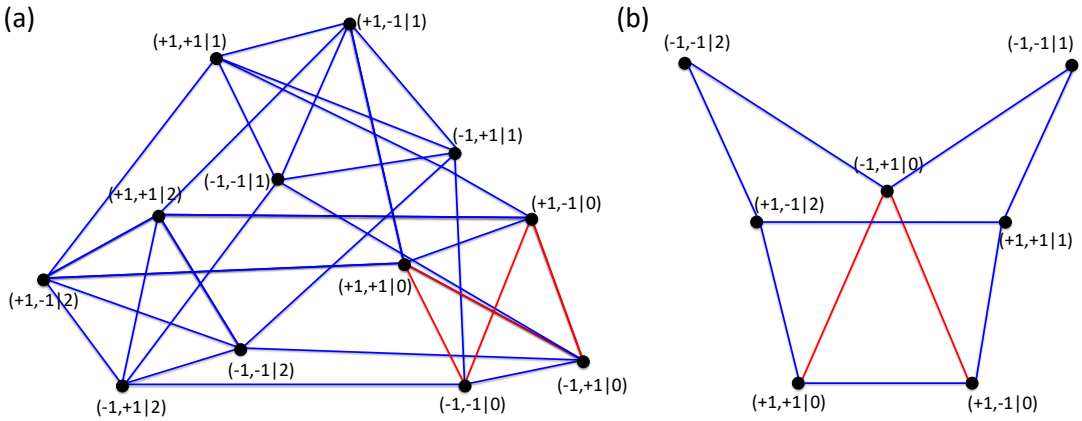


FIG. 6: **(a)** Exclusivity graph of the instrumental Wigner's friend scenario where the red edges are those connecting events involved in the AOE assumption, while the blue ones are standard exclusivity edges. **(b)** Induced subgraph involving the events involved in the genuine facet inequality in Eq. (5).

us to consider the singlet state $|\Psi\rangle = (|01\rangle - |10\rangle)/\sqrt{2}$ and the observables $\hat{A}_0 = -\sigma_z$, $\hat{A}_1 = (\sigma_z - \sigma_x)/\sqrt{2}$, $\hat{A}_2 = (\sigma_z + \sigma_x)/\sqrt{2}$ for Alice, and $\hat{B}_{+1} = \sigma_x$, $\hat{B}_{-1} = -\sigma_z$ for Bob.

To demonstrate this experimentally, polarization-entangled photon pairs from a Spontaneous Parametric Down-Conversion (SPDC) source are used (refer to Sec. IV). Such sources are affected by two distinct kinds of noise [96, 97], which leads to the adoption of the following noise model:

$$\rho = v \cdot |\Psi\rangle\langle\Psi| + \\ + (1-v)\left[\lambda \frac{|HV\rangle\langle HV| + |VH\rangle\langle VH|}{2} + (1-\lambda)\frac{\mathbb{I}}{4}\right], \tag{F1}$$

Here, $|\Psi\rangle = (|HV\rangle - |VH\rangle)/\sqrt{2}$ is the pure singlet state that is ideally being produced by the source. The *white noise* is responsible for the appearance of the maximally mixed state, which corresponds to an overall depolarization of the state. The *coloured noise* is associated with depolarization in a specific direction, and due to reasons related to the generation of the $|HV\rangle$ and $|VH\rangle$ states by SPDC, this noise presents itself as a decoherence term, affecting the specific off-diagonal components of the density matrix. The *visibility* $0 \leq v \leq 1$ parametrizes the portion of the quantum state that is ideal and noiseless, and the *noise proportion* given by $0 \leq \lambda \leq 1$ is the colored noise fraction of the non-singlet mixture.

The parameters v and λ enable us to calibrate the constraints of this quantum strategy, and tuning their values also leads to an identification of the regime in which it is robust to this category of noise. Having fixed the singlet state and measurement operators to the ones mentioned before, we now consider the behavior of Eq. (5) as a function of these noise parameters (refer to Fig. 7). Note the (black) contour that marks the boundary between violation and satisfaction of the inequality; it demarcates the regions and provides the limits of our quantum strategy: the lower bound on visibility goes from about 82.84% when $\lambda = 0$, to about 41.42% when $\lambda = 1$. The dashed contours in red and blue represent results obtained from experimental data.

It is important to highlight that this analysis is only gauging the robustness of the noise created by the SPDC source and quantifying how far the experimentally generated state can be from the ideal scenario and still constitute a violation of the instrumental inequality. More precisely, this calibrates the limit on the error of the state generation, and there are other sources of error subsequently that will lead to a less-than-optimal result; for instance, the switching error of the optical switch (see discussion in Appendix G).

Appendix G: Experimental details and outcome assignment strategy

The instrumental scenario is implemented in the photonic platform already described in Fig. 2(a); here, we specify the assignment of the measurement outcomes in Fig. 8, for both $x = 0$ and $x \neq 0$.

Charlie always performs a σ_z measurement, implemented by a beam displacer that couples polarization and path. When Alice decides to ask her friend the result of his measurement ($x = 0$), a motorized mirror is inserted, and the polarization of each photon is measured through its path, as shown in Fig. 8(a). Therefore, we assign the outcome -1 to the detector placed in front of the motorized mirror, which collects photons from the path corresponding to vertical polarization; consistently, we assign the outcome $+1$ to both detectors after Alice's PBS, which together collect photons coming from the other path. On the other hand, when $x \neq 0$, the friend's interaction with the system is reversed, and Alice's measurement is implemented in the polarization degree of

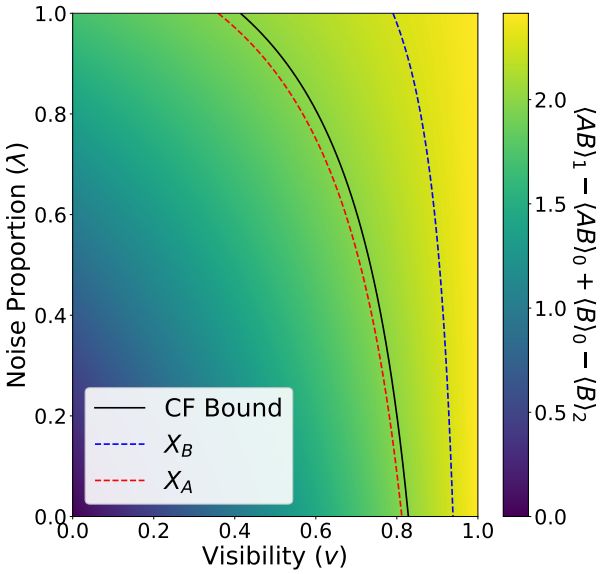


FIG. 7: **Noise robustness of the genuine CF inequality.** The L.H.S of Eq. (5) has been calculated at 250,000 points and plotted as a heatmap by using the noisy model in Eq. (F1) and the optimal measurement settings. The axes span values of the visibility (v) and the noise proportion (λ). The black contour marks the boundary between points that violate and satisfy the genuine CF inequality; the points to the right of the contour populate the region that violates Eq. (5) and is robust to noise. We also show the contours corresponding to the values of the L.H.S. of Eq. (5) from the experimental data (refer to Fig. 3): in blue is the point X_B that corresponds to the experimental violation of the genuine CF facet, and in red is the point X_A, which has a value that lies outside of the CCM polytope but inside the CF one. It is possible to move from one regime to the other, for instance, by changing the parameters of the noise in the state in Eq. (F1): this is the strategy we implemented experimentally by enlarging the coincidence window to measure the point X_A.

freedom by an HWP and a PBS. The path of the photons for these measurement choices is shown in Fig. 8(b). In accordance with the previous case, the outcome $+1$ is again assigned to the detector collecting the photons transmitted by the PBS, which provides the feed-forward for Bob's measurement settings; Alice's remaining detectors are then assigned the outcome -1. In contrast, Bob's measurement station consists of two static polarization measurements; the outcomes' labels are shown in both Fig. 8(a) and Fig. 8(b). The feed-forward from Alice's outcome to Bob's measurement choice is crucial in the instrumental scenario, and is experimentally implemented through an optical switch connected to the detector corresponding to Alice's outcome 1. The employed optical switch is not ideal, meaning that 5% to 15% of the time (depending on the input rate), it fails to direct the incoming photon to the measurement setting indicated by Alice's feed-forward; in these cases, Bob performs a measurement different from the desired one, so reducing the violation of the CF inequality.

Appendix H: Relaxed tracking and changes in CF inequality bounds

The assumption in Eq. (2) certifies the consistency of Charlie's outcome c with Alice's outcome a when $x = 0$. This is called the tracking assumption and is one of the defining features of the CF correlation set:

$$P(a|c, x = 0) = \delta_{a,c}. \tag{H1}$$

We can consider relaxations of the condition Eq. (H1) by allowing non-perfect correlation between a and c bounded up to some value ϵ thus requiring instead that:

$$\begin{aligned} P(a|x = 0, c) &\geq 1 - \epsilon \quad \text{when} \quad a = c \\ P(a|x = 0, c) &\leq \epsilon \qquad \text{when} \quad a \neq c. \end{aligned} \tag{H2}$$

Here ϵ is a measure of the inconsistency of Alice's and Charlie's outcomes: for instance, $\epsilon = 0$ recovers the usual condition of Eq.(H1), while $\epsilon = 1/2$ implies that their outcomes can be completely uncorrelated. In our analysis, having $\epsilon > 0$ generally allows for the possibility of having a nonclassical channel connecting the source with Alice, also when $X = 0$. Adopting this assumption leads to a set of ϵ-Relaxed Nonclassical Causal Friendliness ($\epsilon-$RCF) correlations, of which CF correlations are clearly a subset.

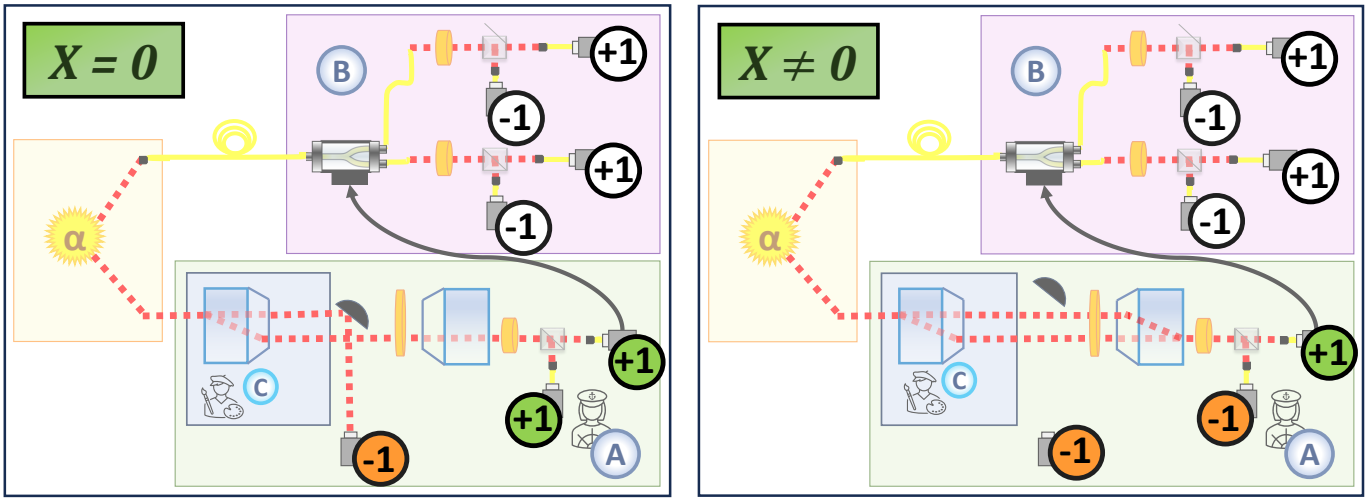
It is interesting to see how well a violation of the CF bound compares with this larger set of $\epsilon-$RCF correlations; among other things, it allows us to relax the requirement that Charlie communicates his outcome to Alice with perfect consistency. Knowing how far we are from the $\epsilon-$modified CF inequality also gives us an upper bound on the $\epsilon-$robustness of the experimental results, that is, an estimate of the maximal deviation of the tracking constraint in Eq. (2) permissible to warrant a genuine violation of the $\epsilon-$RCF inequality. We will show that when ϵ is small, the shift in the bound of the CF expression

$$I_{\text{inst}} = \langle AB\rangle_1 - \langle AB\rangle_0 + \langle B\rangle_0 - \langle B\rangle_2 \overset{CF}{\leq} 2 \tag{H3}$$

is small as well, in the sense of the theorem below:

Theorem 3. *For $\epsilon \leq 1/2$, a tight upper bound of I_{inst} for ϵ-RCF behaviours is*

$$I_{\text{inst}} \overset{\epsilon-\text{RCF}}{\leq} 2 + 2\epsilon. \tag{H4}$$



(a) (b)

FIG. 8: **Assignment of outcomes in the experimental setup.** Bob's measurement station performs fixed polarization measurements based on Alice's outcome, and the outcome labels are therefore fixed. By contrast, the outcome labels in Alice's station depend on the measurement choice x, which determines whether Alice measures the photon's path (corresponding to asking the friend for their outcome) or its polarization (corresponding to reversing the interaction and then performing a measurement). The corresponding labeling convention is illustrated in panel **(a)** for $x = 0$ and in panel **(b)** for $x \neq 0$.

First, let us mention that measures of "non-absoluteness" in the context of Local Friendliness constraints over the Bell-scenario-based extensions of the Wigner's friend scenario (sometimes termed canonical Local Friendliness-scenarios [16]) were studied in detail in Ref. [73]. There, direct $\epsilon-$relaxations of the condition in Eq. (H1) and their implications for the set of admissible correlations were also explored. However, the relaxations they considered were of the form (in our language)

$$P(a = c|x = 0) \geq 1 - \epsilon. \tag{H5}$$

This form of relaxation is generally *not equivalent* to the form we require in Eq. (H2), which explicitly refers to the conditional marginals $P(a|x, c)$ (instead of the joint $p(a, c|x)$), though both recover the usual criterion when $\epsilon = 0$. Indeed, the form of relaxation we consider in Eq. (H2) is more closely related to allowing some local degree of indeterminism, or more precisely nonclassicality, as was done by Hall [98] and, explicitly, the notion of predetermination (more generally, classicality) up to error ϵ for subsets of settings in the supplemental material (pages 19-20) of Ref. [99]. Emphatically, however, those results were in the context of Bell inequalities (and different physical motivations), in contrast to the instrumental-based scenario we consider here.

We shall next demonstrate a proof of Theorem 3. Our derivation is greatly inspired by that of Ref. [98] and we adopt similar notation as was used in the appendix of Ref. [73]. First, let us observe that owing to the linearity in probabilities of the expression I_{inst}, and the fact that $\wp(a, b|x) = \sum_c P(a, b, c|x) = \sum_c P(a, b|x, c)P(c)$ it is possible to equivalently write $I_{\text{inst}} = \sum_c P(c) I^c_{\text{inst}}$, where I^c_{inst} denotes an expression of similar form evaluated on the conditional distributions $P(a, b|x, c)$. We will derive a bound on I^c_{inst} that holds for all c, hence by averaging over c also for I_{inst}.

By writing

$$\begin{aligned} P(+1, +1|x, c) &= \gamma^c_x \\ P(a = +1|x, c) &= \alpha^c_x \\ P(b = +1|x, c) &= \beta^c_x \end{aligned} \tag{H6}$$

and using the fact that $\alpha^c_x = \sum_b P(+1, b|x, c)$ and so on, the full probability distributions $P(a, b|x, c)$ can be written as

$$\begin{aligned} P(+1, +1|x, c) &= \gamma^c_x \\ P(+1, -1|x, c) &= \alpha^c_x - \gamma^c_x \\ P(-1, +1|x, c) &= \beta^c_x - \gamma^c_x \\ P(-1, -1|x, c) &= 1 - \beta^c_x - \alpha^c_x + \gamma^c_x \, . \end{aligned} \tag{H7}$$

Using these expressions, I^c_{inst} takes the form

$$I^c_{\text{inst}} = 4(\gamma^c_1 - \gamma^c_0) + 2(\alpha^c_0 - \alpha^c_1 - \beta^c_1 - \beta^c_2) + 4\beta^c_0 \, . \tag{H8}$$

Now, for all probability distributions, we always have

$$\max\{0, \alpha^c_x + \beta^c_x - 1\} \leq \gamma^c_x \leq \min\{\alpha^c_x, \beta^c_x\} \, , \tag{H9}$$

implying the upper bound

$$\begin{aligned} I^c_{\text{inst}} \leq & 4 \min\{\alpha^c_1, \beta^c_1\} - 4 \max\{0, \alpha^c_0 + \beta^c_0 - 1\} \\ & + 4\beta^c_0 + 2(\alpha^c_0 - \alpha^c_1 - \beta^c_1 - \beta^c_2). \end{aligned} \tag{H10}$$

Now note that if $\alpha_0^c + \beta_0^c - 1 \leq 0$ then $-\max\{0, \alpha_0^c + \beta_0^c - 1\} + \beta_0^c = \beta_0^c \leq 1 - \alpha_0^c$ and if $\alpha_0^c + \beta_0^c - 1 > 0$ then $-\max\{0, \alpha_0^c + \beta_0^c - 1\} + \beta_0^c = 1 - \alpha_0^c$ so that in either case

$$\begin{aligned} I_{\text{inst}}^c &\leq 4\min\{\alpha_1^c, \beta_1^c\} + 4 - 4\alpha_0^c + 2(\alpha_0^c - \alpha_1^c - \beta_1^c - \beta_2^c) \\ &= 4 + 4\min\{\alpha_1^c, \beta_1^c\} - 2(\alpha_0^c + \alpha_1^c + \beta_1^c + \beta_2^c). \end{aligned} \tag{H11}$$

Recalling that $2\min\{x, y\} = x + y - |x - y|$ the above means that

$$\begin{aligned} I_{\text{inst}}^c &\leq 4 + 2(\alpha_1^c + \beta_1^c - |\alpha_1^c - \beta_1^c|) - 2(\alpha_0^c + \alpha_1^c + \beta_1^c + \beta_2^c) \\ &= 4 - 2|\alpha_1^c - \beta_1^c| - 2(\alpha_0^c + \beta_2^c) \\ &\leq 4 - 2\alpha_0^c \,. \end{aligned} \tag{H12}$$

The upper bound $I_{\text{inst}}^c \leq 4 - 2\alpha_0^c$ is nontrivial, and recovers the bound $I_{\text{inst}}^c \leq 2$ immediately if $\alpha_0^c = 1$. Note that by assumption, this is true for CF correlations for the case $c = +1$; however, the derivation of this bound did not require the use of other causal constraints and hence the bound $I_{\text{inst}}^c \leq 2$ is seen to be valid for all distributions with $\alpha_0^c = 1$ independently of whether the $P(a, b|x, c)$ satisfy the instrumental inequalities in particular or not. The bound is also seen to be close to 2 for values of α_0^c close to 1, which makes it suitable to investigate ϵ-RCF. However, for $\alpha_0^c = 0$ (and values close to 0), which is implied by the CF assumptions for $c = -1$, it gives no nontrivial constraints (recovering the algebraic upper bound). It furthermore seems that owing to the lack of symmetry in the expression I_{inst}^c, the same derivation with $\alpha_0'^c = (1 - \alpha_0^c)$, which would imply the other case, does not lead to a similar conclusion. Fortunately, we can cover the case $\alpha_0^c \simeq 0$ by a different approach.

Recall the instrumental inequalities, which are valid for all correlations compatible with the generalized causal Markov condition:

$$P(a, +1|x, c) + P(a, -1|x', c) \leq 1 \;\; \forall a, (x, x') \,. \tag{H13}$$

Using Eqs. (H7), these reduce to the following forms of inequality constraints

$$\begin{aligned} \gamma_x^c + \alpha_{x'}^c - \gamma_{x'}^c &\leq 1 \\ \beta_x^c - \gamma_x^c - \beta_{x'}^c - \alpha_{x'}^c + \gamma_{x'}^c &\leq 0 \end{aligned} \tag{H14}$$

for all pairs x, x'. In particular

$$\gamma_1^c + \alpha_2^c - \gamma_2^c \leq 1 \tag{H15}$$

$$\beta_0^c - \gamma_0^c - \beta_1^c - \alpha_1^c + \gamma_1^c \leq 0 \tag{H16}$$

$$\beta_0^c - \gamma_0^c - \beta_2^c - \alpha_2^c + \gamma_2^c \leq 0 \tag{H17}$$

and by adding the above together one gets

$$2(\gamma_1^c - \gamma_0^c) \leq 1 + \alpha_1^c + \beta_1^c + \beta_2^c - 2\beta_0^c \,. \tag{H18}$$

Substituting above to Eq. (H8) yields the upper bound

$$I_{\text{inst}}^c \leq 2 + 2\alpha_0^c \,. \tag{H19}$$

We see that now for $\alpha_0^c = 0$, which by CF assumption holds for the case $c = -1$, the upper bound $I_{\text{inst}}^c \leq 2$ is again recovered. However, the nontrivial bound in Eq. (H19) holds for all (conditional) correlations obeying the instrumental inequalities, not just those which satisfy CF. Together with (H13) they give easily obtainable bounds on ϵ-RCF. Indeed note first that for $\alpha_0^c \in [0, 1/2]$ the bound in Eq. (H19) is more constraining than Eq. (H13), whereas the opposite holds true for $\alpha_0^c \in [1/2, 1]$. The bounds coincide and allow for the maximum value of I_{inst}^c when $\alpha_0^c = 1/2$, so that by averaging over c the maximum of I_{inst} is seen to be at most 3, which is already less than the algebraic upper bound. On the other hand, by definition of Eq. (H2) $\alpha_0^c \in [0, \epsilon] \cup [1 - \epsilon, 1]$ for all c, corresponding to the cases $P(a = +1|x = 0, c = +1) \geq 1 - \epsilon$ and $P(a = +1|x = 0, c = -1) \leq \epsilon$, where the latter follows from normalization. Consequently for $\alpha_0^c \in [1 - \epsilon, 1]$ with $\epsilon \leq 1/2$ the bound in Eq. (H13) is more constraining, and gives $I_{\text{inst}}^c \leq 4 - 2(1 - \epsilon) = 2 + 2\epsilon$ whereas for $\alpha_0^c \in [0, \epsilon]$ with $\epsilon \leq 1/2$ Eq. (H19) is tighter and likewise gives $I_{\text{inst}}^c \leq 2 + 2\epsilon$. Thus, by averaging over c, the upper bound

$$I_{\text{inst}} \leq 2 + 2\epsilon \tag{H20}$$

is obtained for all ϵ-RCF correlations with $\epsilon \leq 1/2$.

It remains to show that this bound is achievable by an ϵ-RCF model. For this purpose, consider a model which has $\alpha_0^{-1} = \epsilon$ and $P(c = -1) = 1$ so that $I_{\text{inst}} = I_{\text{inst}}^{c=-1}$ and by comparison of Eqs. (H8) and (H18) it is seen that the equality $I_{\text{inst}}^{c=-1} = 2 + 2\epsilon$ holds if Eq. (H18) holds with equality. Additionally, we have to check that: 1) the value assignments satisfy the instrumental inequalities (H14)–(H15), 2) the assignments are consistent with the relation to the probability distributions $P(a, b|x, c)$ as in Eqs. (H7). Consider (see acknowledgements) the set:

$$\begin{array}{lll} \gamma_0^{-1} = 0 & \alpha_0^{-1} = \epsilon & \beta_0^{-1} = 1 - \epsilon \\ \gamma_1^{-1} = \epsilon & \alpha_1^{-1} = \epsilon & \beta_1^{-1} = 1 - \epsilon \\ \gamma_2^{-1} = 0 & \alpha_2^{-1} = 1 - \epsilon & \beta_2^{-1} = 0 \,. \end{array} \tag{H21}$$

We can verify that the assignment in Eq. (H21) is indeed valid by going through all variations of Eqs. (H14)–(H15) (with $c = -1$). By writing $f(x, x') = \gamma_x^{-1} + \alpha_{x'}^{-1} - \gamma_{x'}^{-1}$ into a vector $\vec{f}$ and $g(x, x') = \beta_x^{-1} - \gamma_x^{-1} - \beta_{x'}^{-1} - \alpha_{x'}^{-1} + \gamma_{x'}^{-1}$ into a vector $\vec{g}$ with components ordered by $(x, x') = (0, 1), (0, 2), (1, 0), (1, 2), (2, 0), (2, 1)$. After direct computation, one obtains

$$\vec{f} = [0, 1 - \epsilon, 2\epsilon, 1, \epsilon, 0]^T \tag{H22}$$

$$\vec{g} = [0, 0, -2\epsilon, -\epsilon, -1, -1 + \epsilon]^T \,. \tag{H23}$$

Both families of inequalities, which demand $f(x, x') \leq 1$ and $g(x, x') \leq 0$ for all x, x' respectively, are seen to be satisfied for all $\epsilon \in [0, 1/2]$. Furthermore, since Eq. (H9) holds for all x, we confirm that the solution in Eq. (H21) indeed specifies a valid $\epsilon-$parametrized family of behaviours $\wp^\epsilon(a, b|x)$ which satisfy $\epsilon-$RCF constraints and

reach the upper bound $I_{\text{inst}} = 2 + 2\epsilon$. For completeness, we write the behaviour explicitly. In this model $\wp^{\epsilon}(a,b|x) = \sum_c P^{\epsilon}(a,b|x,c)P(c) = P^{\epsilon}(a,b|x,c=-1)$ and where $P^{\epsilon}(a,b|x,c=-1)$ is obtained using the solution in Eq. (H21) from Eqs.(H7) as

$$
\begin{array}{ll}
P(+1,+1|x=0,c=-1)=0 & P(+1,-1|x=0,c=-1)=\epsilon \\
P(-1,+1|x=0,c=-1)=1-\epsilon & P(-1,-1|x=0,c=-1)=0 \\
\hline
P(+1,+1|x=1,c=-1)=\epsilon & P(+1,-1|x=1,c=-1)=0 \\
P(-1,+1|x=1,c=-1)=1-2\epsilon & P(-1,-1|x=1,c=-1)=\epsilon \\
\hline
P(+1,+1|x=2,c=-1)=0 & P(+1,-1|x=2,c=-1)=1-\epsilon \\
P(-1,+1|x=2,c=-1)=0 & P(-1,-1|x=2,c=-1)=\epsilon
\end{array}
\tag{H24}
$$