The Henryk Niewodniczański
Institute of Nuclear Physics
Polish Academy of Sciences

Ewa Borsuk

Entanglement generation in the no-touching scenario: causal aspects & explicit protocols

Supervisor:
dr hab. Paweł Błasiak

Thesis submitted for Degree of
Doctor of Philosophy in Physics

Kraków, 2023

Acknowledgements

I want to deeply thank my supervisor, Paweł Błasiak, for all the helpful discussions, and guidance and for showing me so many inspiring aspects of quantum mechanics and causality. I also want to thank Marcin Markiewicz for his helpful comments. My gratitude also goes to Ehtibar N. Dzhafarov for all insightful discussions about quantum contextuality and to Jacob Barandes for organising excellent seminars on quantum foundations.

I also want to send my gratitude towards my colleagues, Timothée Hoffreumon, Diana Taschetto and Carlo Marconi, for the amazing exchange of viewpoints on quantum foundations and the usage of complex numbers in quantum mechanics. I also want to thank my friends from PhD studies at KISD, especially Wiktoria Stańczyk and Nadine Hammoud, for their support and good company in good and bad times.

Lastly, I want to thank my husband, my mom, my family and close friends for all the support and love. I wouldn't make it without you. Thank you for your presence.

Abstract

The central theme of this dissertation is entanglement, its optical generation schemes and the ensuing causal aspects related to quantum foundations. It is divided into two parts following the natural order of increasing generality. The first part focuses on practical schemes of entanglement generation in passive linear optical designs within the so-called *no-touching* paradigm. In the second part, some fundamental topics are discussed related to the problem of *post-selection* in entanglement generation schemes, as well as the meaning and quantitative assessment of the violation of Bell inequalities within the causal approach.

The first part starts with an introduction to the concept of *no-touching* in state generation protocols. We are interested in the following question: *What is the range of entangled states that can be obtained in such schemes?* To illustrate the generality of the proposal, we construct explicit protocols that generate an arbitrary state of three qubits and a multipartite W state for an arbitrary number of qubits, which work for both bosons and fermions. We study efficiency of these protocols and compare them with other proposals in the literature to show the advantages of the *no-touching* designs.

In the second part, we turn to the famous Bell experiment, revisiting some fundamental causal concepts behind it. We first observe that oftentimes entanglement generating protocols (including the *no-touching* proposal) involve post-selection which opens the door for a *selection bias*. Here, we give an illustrative example of the quantum three box paradox where all the mystery can be merely explained as an effect of post-selection which is responsible for the emergence of non-causal correlations. This leads to the question: *Can those protocols be actually used for conclusive tests of Bell inequalities in the presence of post-selection?* We formulate a simple criterion, the so-called *all-but-one principle*, which is proved to safeguard the causal conclusions from the violation of Bell inequalities that are drawn using post-selected data. It is applicable to a broad range of experimental protocols (including the *no-touching* proposal).

The second part ends with an unbiased discussion of the assumptions in Bell experiments. We aim at the question: *What is the causal cost of violation of Bell inequalities, and hence how do the causal assumptions compare one to another?* Our figure of merit is the natural concept of the so-called *fractional measure* which ask about the frequency of violation of a given assumption (that is, the presence of certain causal arrows). We show that such defined measures of *locality* and *free choice* are the same for Bell-type scenarios. Furthermore, we explicitly calculate those measures for the standard bipartite, two-setting and two-outcome Bell scenario. Remarkably, this result gives a simple and natural quantitative *interpretation* of the amount of violation of the CHSH inequalities. We also discuss in detail the case of quantum mechanical statistics.

All the results were published in five journal articles, which form a collection of papers submitted for the doctoral degree.

Streszczenie

Tematyka niniejszej rozprawy dotyczy splątania kwantowego, generowania stanów w schematach optycznych oraz zagadnień związanych z przyczynowością w podstawach mechaniki kwantowej. Praca jest podzielona na dwie części zgodnie z rosnącą ogólnością przedstawionych zagadnień. Pierwsza część koncentruje się na praktycznych schematach generowania splątania w pasywnych liniowych układach optycznych w ramach tak zwanego paradygmatu *bez kontaktu*. W drugiej części zostają omówione wybrane zagadnienia związane z *post-selekcją* używaną w schematach generowania stanów splątanych, oraz znaczeniem i ilościową oceną łamania nierówności Bella w ramach podejścia przyczynowego.

Pierwsza część rozprawy rozpoczyna się wprowadzeniem do *bezkontaktowych* protokołów generowania stanów. Dyskusji zostaje poddane pytanie: *Jaki jest zakres stanów splątanych, który można uzyskać w takich protokołach?* Aby odpowiedzieć na to pytanie, konstruujemy optyczne protokoły generujące dowolny stan trzech qubitów oraz stany W dla dowolnej liczby qubitów. Przedstawione schematy działają zarówno dla bozonów jak i fermionów. Badamy również efektywność tych protokołów i porównujemy je z innymi znanymi schematami w celu przeanalizowania zalet paradygmatu *bez kontaku*.

W drugiej części rozprawy analizujemy słynny eksperyment Bella, powracając do fundamentalnych problemów związanych z przyczynowością. Przyczynkiem do tych rozważań jest obserwacja, że protokoły generowania splątania (w tym również schemat *bez kontaktu*) często opierają się o tzw. *post-selekcję*, która może prowadzić do błędnych konkluzji i wprowadzenia nadmiarowych korelacji. Aby zilustrować problem, analizujemy przykład kwantowego paradoksu trzech pudełek, w którym post-selekcja może odpowiadać za powstanie paradoksu przez wprowadzenie nieprzyczynowych korelacji. Prowadzi to do pytania: *Czy bezkontaktowe protokoły mogą być wykorzystane do konkluzywnych testów nierówności Bella w obecności post-selekcji?* W odpowiedzi na tak zadane pytanie, formułujemy proste kryterium, tak zwaną zasadę *wszystko minus jeden* i pokazujemy, że potwierdza ona wiarygodność wniosków wyciągniętych z łamania nierówności Bella nawet przy użyciu danych z post-selekcji. Zasada ta ma zastosowanie dla szerokiej gamy protokołów eksperymentalnych (w tym również schematów *bez kontaktu*).

Druga część kończy się dyskusją na temat założeń eksperymentów Bella. Naszym celem jest pytanie: *Jaki jest koszt łamania nierówności Bella, oraz jak porównać ich przyczynowe założenia?* Używamy pojęcia tzw. *miar ułamkowych*, które opisują częstotliwość naruszenia danego założenia. Pokazujemy, że tak zdefiniowane miary *lokalności* i *wolnego wyboru* dla eksperymentów typu Bella są takie same. Dodatkowo, jawnie obliczamy wartości tych miar dla standardowego eksperymentu Bella z dwoma eksperymentatorami, dwoma ustawieniami aparatury i dwoma wynikami. Co ciekawe, ten rezultat daje prostą i naturalną ilościową *interpretację* stopnia łamania nierówności CHSH. Omawiamy również szczegółowo przypadek statystyki w mechanice kwantowej.

Wszystkie wyniki zostały opublikowane w pięciu artykułach naukowych, które stanowią zbiór prac przedłożonych w ramach niniejszej rozprawy.

Contents

1 Introduction

Quantum entanglement is a new quality that is believed to be responsible for technological breakthroughs in the coming quantum information era. It has become possible due to the departure from the classical description on the microscopic level, which has consequences on the observed macroscopic level. This comes not without tension with our most cherished intuitions about the world. In his seminal work, John S. Bell pointed out that certain natural assumptions about how cause-and-effect works on the fundamental level contradict quantum predictions. Those results were later experimentally confirmed and recognised with the Nobel Prize in Physics 2022 awarded to A. Aspect, J. F. Clauser and A. Zeilinger "for experiments with entangled photons, establishing the violation of Bell inequalities and pioneering quantum information science".

My research interests in this thesis concern certain topics in quantum entanglement and its consequences for the causal picture of the world. This is a theoretical work motivated by both practical applications and fundamental questions. More specifically, I am focused on the problem of

i) entanglement generation in passive linear optical designs (within the so-called *no-touching* paradigm),

ii) its use for the purpose of fundamental tests, and

iii) the causal meaning of the observed violation of Bell inequalities.

Formally, my dissertation is a collection of the following five thematically connected research articles:

[1] P. Blasiak, E. Borsuk and M. Markiewicz.
Arbitrary entanglement of three qubits via linear optics.
Scientific Reports (Nature) **12** 21596 (2022).

[2] P. Blasiak, E. Borsuk, M. Markiewicz and Y.-S. Kim.
Efficient linear optical generation of a multipartite W state.
Physical Review A **104** 023701 (2021).

[3] P. Blasiak and E. Borsuk.
Causal reappraisal of the quantum three-box paradox.
Physical Review A **105** 012207 (2022).

[4] P. Blasiak, E. Borsuk, and M. Markiewicz.
On safe post-selection for Bell nonlocality: Causal diagram approach.
Quantum **5** 575 (2021).

[5] P. Blasiak, E. M. Pothos, J. M. Yearsley, C. Gallus, and E. Borsuk.
Violations of locality and free choice are equivalent resources in Bell experiments.
Proceedings of the National Academy of Sciences (PNAS) **118** e2020569118 (2021).

EurekAlert! (Press note May 2021):
We know the cost of free choice and locality - in physics and not only.

Those articles provide answers to three research problems tackled during my Doctoral studies. Here is the skeleton of three research questions around which this dissertation is built (following the natural order of increasing generality and foundational interest).

Q1. What is the range of entangled states that can be generated in the *no-touching* scenario? What is the efficiency of those protocols? See articles [1, 2].

Q2. Can those protocols be used for conclusive tests of Bell inequalities in the presence of post-selection? See articles [3, 4].

Q3. What is the causal cost of violation of Bell inequalities, and hence how do the causal assumptions compare one to another? See article [5].

In the following, I elaborate on those research problems and explain the results obtained in the articles [1–5]. Accordingly, **Section 2.1** gives a gentle introduction to the *no-touching* state generation scenarios. In **Section 2.2**, I present two explicit protocols illustrating the generality and practical utility of the proposed scheme (answering **Q1**): the first one concerns the generation of an arbitrary state of three qubits [1], the second one discusses the generation of the multipartite W state for an arbitrary number of qubits [2]. Then I proceeded to the foundational questions regarding entanglement. **Section 3.1** sketches out the framework of Bell experiments and gives a brief discussion of its causal implications. This raises the issue of post-selection in entanglement generation schemes (explicitly present in the *no-touching* scenarios), which is discussed in **Section 3.2**. In **Section 3.2.1**, I illustrate the warning of a selection bias by a careful causal analysis of the quantum three-box paradox, where post-selection can explain paradoxical correlations in the experiment [3]. In order to allay concerns of misinterpreting the violation of Bell inequalities in the presence of post-selection, I describe in **Section 3.2.2** the so-called *all-but-one* principle [4], which protects the causal conclusions against the selection bias (answering **Q2**). Finally, in **Section 3.3**, I approach the problem of a quantitative comparison and explicit calculation of the cost of violation of the locality and free choice assumptions in Bell inequalities, using the intuitive concept called *fractional measure* [5] (answering **Q3**). **Section 4** summarises all the results.

2 Entanglement generation in the *no-touching* scenario

Entanglement is considered a novel resource in a number of quantum information tasks [6]. To name a few, it is used to provide security in quantum cryptographic protocols [7], speeds up quantum algorithms [8] or promises better performance in quantum metrology [9]. With those practical applications comes a nagging question of where those advantages come from and what this "new" quality, called entanglement, tells us about the nature of reality [10–13]. This question belongs to the vast domain of quantum foundations; see **Section 3** for a brief discussion.

Since entanglement is a key quantum resource, both for practical applications and fundamental research, it is essential to have it on demand in various forms. The broader the range of entangled states and methods of their generation, the better (so that they can be used for a particular task at hand). However, it is fair to say that there is no systematic and practical methodology for generating arbitrary entangled quantum states, be it multipartite or higher dimensional entanglement. Each type of physical implementation of a qubit (photon polarisation, electronic spin, dual-rail encoding, atomic energy levels, electric current, etc.) is subject to its own generation and manipulation techniques. See Refs. [14–16] for a review of modern optical methods of state generation. Typically one needs to resort to non-linearities and/or post-selection. Another serious difficulty lies in the complexity of the space of multiparticle states. It is instructive to compare the single particle case [17] where all states are accessible by passive linear optical means, and the multi-particle case where the state space splits into a continuous myriad of classes under passive linear optical transformations [18] (whose complete classification is currently beyond reach). Notably, some theoretical progress can be made in the special case where it is known that the so-called stochastic local operations and classical communication (SLOCCs) splits the state space of three qubits into two classes of genuinely entangled states (of the GHZ-state and the W-state type) [19].

In the first part of my dissertation, I deal with entanglement generation schemes called the *no-touching* scenarios. This method relies on the intrinsic indistinguishability of particles and quantum interference. It aims at the generation of multiparticle entangled states using passive linear optics with post-selection. This method has been recently proposed in Ref. [20] and originates in the seminal works of Yurke-Stoler [21, 22]. **Section 2.1** briefly introduces the *no-touching* scenario and gives two illustrative examples. It is natural to ask about the generality and efficiency of the proposed scheme. I worked out two nontrivial examples of state generation. In **Section 2.2.1**, I describe the explicit *no-touching* scheme of generating an arbitrary state of three qubits, which works with constant efficiency for any desired state (based on Ref. [1]). Then, in **Section 2.2.2**, I ask about the performance of the *no-touching* scenario for an arbitrary number of qubits on the prominent example of the multipartite W state and compare it with the best-known protocols (based on Ref. [2]).

2.1 No-touching scenario: General scheme

The original idea derives from the proposal of the Yurke-Stoler interferometer [21], where Bell states are produced in a no-touching way. Later, the authors developed a more advanced protocol for generating the GHZ state for three particles [22]. In Ref. [20], the idea of Yurke-Stoler was revisited, and the general *no-touching* protocol for entanglement generation was proposed. See also Refs. [23–26] for some ideas along those lines.

The key idea behind the *no-touching* scenario is an observation that particles do not have to touch (or meet) one another in order to create an entangled state. This contradicts the common intuition that entanglement is generated due to some "interaction" process, which requires particles to be present in the same location at the same time (i.e., informally speaking, touch one another). From the dynamical point of view, it boils down to the presence of mixing terms in the Hamiltonian coupling the respective modes of the system (cf., the account of fundamental interactions in particle physics or entanglement generation via spontaneous parametric down-conversion in quantum optics). This intuition also prevails in the kinematical picture of quantum interference on a beam splitter, where bosons show the bunching effect and fermions due to the Pauli exclusion principle show the anti-bunching effect [14, 27]. [Here, the correlation between the impinging particles arises as a consequence of particle statistics and the commutation relations at the touching point]. So, the term "*no-touching*" refers to situations where the particles do not touch/meet one another during the experimental protocol.

Here, we are interested in entanglement generation in (passive) linear optics. That is, we consider a set of optical modes paths (or modes) forming a circuit built of mirrors, phase shifters and beam splitters. The *no-touching* scenario starts with a given number of independent particles injected in the input and asks for a given type of entanglement generated at the output (typically in the so-called dual-rail encoding). Those protocols are subject to certain constraints in design (due to the no-touching requirement) and allow post-selection at the output. It is essential that the particles used in the experiments be indistinguishable/identical, i.e. they obey boson or fermion statistics.

The general setup for the *no-touching* protocol is presented in Fig. 1 (following Ref. [20]). It consists of N optical paths grouped into K input subsystems $A_1, ..., A_K$ and output subsystems $B_1, ..., B_K$. Let the paths in the respective subsystems be labelled as follows: $A_k = \{i_k, i_k + 1, ...\}$ and $B_k = \{j_k, j_k + 1, ...\}$ for $k = 1, ..., K$, where $1 < i_1 < i_2 < ... < i_K < N$ and $1 < j_1 < j_2 < ... < j_K < N$. For the purpose of entanglement encoding (dual-rail), within each output subsystem B_k, we distinguish pairs of paths $\bar{B}_k := \{j_k, j_k + 1\} \subset B_k$. The *no-touching* scenario consists of the following five steps:

1. *Injection* of K independent particles into the system, where each subsystem A_k receives a single particle in the optical path i_k.

2. *Local unitary* transformations $U_1, ..., U_K$ on each subsystem $A_1, ..., A_K$.

3. *Permutation* of optical paths $\sigma \in S_V$.

4. *Local unitary* transformations $V_1, ..., V_K$ on each subsystem $A_1, ..., A_K$.

5. *Post-selection* on the presence of a single particle in each subsystem $\bar{B}_1, ..., \bar{B}_K$.

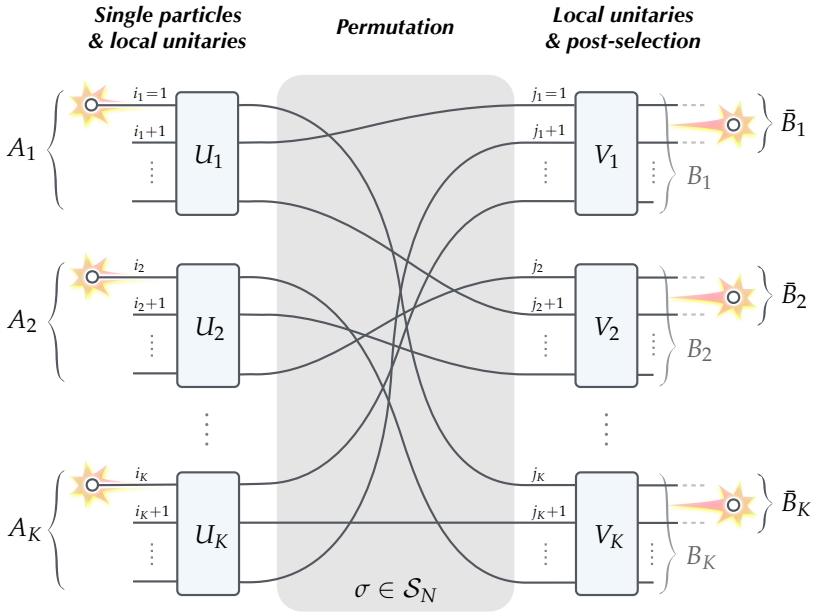


Figure 1: A general scheme of the *no-touching* scenario (following Ref. [20]). The protocol starts with K independent particles injected into paths $j_1, ..., j_K$. Then, the circuit implements a series of local unitary transformations $U_1, ..., U_K$, permutation of optical paths σ and local unitary transformations $V_1, ..., V_K$. At the end, the system is post-selected on a single detection in each dual-rail qubit $\bar{B}_1, ..., \bar{B}_K$.

Schematically, this leads to the evolution of the system described as follows

$$\underbrace{a^\dagger_{i_1} ... a^\dagger_{i_K}}_{\substack{K \text{ independent} \\ \text{particles}}} |0\rangle \xrightarrow{U_1,...,U_K} ... \xrightarrow{\sigma} ... \xrightarrow{V_1,...,V_K} ... \xrightarrow{\text{post-sel.}} \left\{ \begin{array}{c} \text{entangled state in} \\ \text{dual-rail encoded} \\ \text{qubits } \bar{B}_1, ..., \bar{B}_K \end{array} \right\}, \quad (1)$$

where unitaries act on each mode independently $a^\dagger_k \xrightarrow{U} \sum_l U_{kl}\, a^\dagger_l$, permutation switches the modes $a^\dagger_k \xrightarrow{\sigma} a^\dagger_{\sigma(k)}$ and post-selection boils down to retaining in the final expression only the terms with one-and-only-one creation operator in each pair of modes $\bar{B}_k = \{j_k, j_k + 1\}$ for all $k = 1, ..., K$.

Note that it has analogous structure to the Yurke-Stoler interferometer [21, 22]. In the optical context, one can think of the paths as spatially separated modes. The design can be implemented so that subsystems $A_1, ..., A_K$ and $B_1, ..., B_K$ are located in separated space regions. In this sense, the unitaries $U_1, ..., U_K$ and $V_1, ..., V_K$ can be seen as local transformations. Furthermore, the permutation σ can be arranged in a way that the paths do not cross.

In the last step, target qubits are encoded in pairs of paths $\bar{B}_k = \{j_k, j_k + 1\}$. Observe that post-selection on a single particle in each subsystem $\bar{B}_1, ..., \bar{B}_K$ (i.e., events with exactly one particle present in each $\bar{B}_k$) guarantees proper *dual-rail encoding* of qubits. In quantum mechanics, the state of a qubit is described by a superposition of two computational basis states. In the case of dual-rail encoded qubit $\bar{B}_k$, these basis states, denoted as $|\uparrow_k\rangle$ and $|\downarrow_k\rangle$, correspond to the particle being present in the respective path j_k and $j_k + 1$. Thus, using the second-quantized formalism of creation-annihilation operators, a general qubit state encoded in $\bar{B}_k$ takes the form $|\psi\rangle_k = \alpha\,|\uparrow_k\rangle + \beta\,|\downarrow_k\rangle = \left(\alpha\, a^\dagger_{j_k} + \beta\, a^\dagger_{j_k+1}\right)|0\rangle$, where $|0\rangle$ is the vacuum and $|\uparrow_k\rangle = a^\dagger_{j_k}|0\rangle$ (particle in path j_k) and $|\downarrow_k\rangle = a^\dagger_{j_k+1}|0\rangle$ (particle in path $j_k + 1$).

We said that the *no-touching* scenario consists in coincident injection of K independent particles which later on *do not* touch/meet with each other at any stage of the experiment. One can wonder if the scheme presented in Fig. 1 really meets this requirement. Indeed, in Steps 1, 2 and 3, the setup can be arranged so that the particles do not cross their paths (subsystems $A_1, ..., A_K$ can be spatially separated, and the permutation σ may keep the paths apart). The only place where the particles can meet is Step 4, where the paths from different A's merge in subsystems B's. However, those cases are rejected by post-selection in the final Step 5, which rejects the events where the particles could have met in the same subsystem B_k at the end. Hence, the term *'no-touching'* applies to the cases selected in the experiment.

Let us note that in this form the protocol is well-suited for multipartite Bell tests, where the experiment ends by local measurements on each dual-rail qubit $\bar{B}_1, ..., \bar{B}_K$. However, if those qubits are to be used for other complicated quantum information tasks, one needs to ensure proper post-selection before they are further processed. This can be implemented in Step 5 with non-demolition measurements carried out on each qubit $\bar{B}_1, ..., \bar{B}_K$.

In Ref. [20], the *no-touching* scenario was proposed as a novel tool for entanglement generation. It is advertised as demonstrating the utility of particle indistinguishability as a resource of entanglement for practical applications. Furthermore, the proposal uses minimal resources to create entanglement: just passive linear optics and no entanglement in the input (only K independent particles).

There are two natural questions about the proposed scheme. The first concerns its generality and practical utility (due to inherent post-selection).

Q1. *What is the range of entangled states that can be generated in the no-touching scenario? What is the efficiency of those protocols?*

The second question concerns the validity of conclusions drawn from Bell-type experiments with post-selection (due to vulnerability to selection bias).

Q2. *Can those protocols be used for conclusive tests of Bell inequalities in the presence of post-selection?*

Both questions are addressed in **Section 2.2** and **Section 3.2** respectively.

To illustrate the general scheme of state generation described above, we end this section with two simple examples rewriting the original Yurke-Stoler interferometric proposal [21, 22].

2.1.1 Example: Generation of the Bell state

Let us consider the following maximally entangled two-qubit Bell state:

$$|\psi_{Bell}\rangle = \frac{1}{\sqrt{2}}(|\uparrow\uparrow\rangle + |\downarrow\downarrow\rangle), \tag{2}$$

where $\{\uparrow, \downarrow\}$ is a computational basis.

An optical scheme generating the Bell state Eq. (2) in the no-touching scenario is presented in Fig. 2. The protocol starts with the injection of two independent particles into optical paths 1 and 3, one per each subsystem $A_1 = \{1, 2\}$ and $A_2 = \{3, 4\}$. Then, there are two local unitaries described by the Hadamard matrix (implemented by a beam-splitter):

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{3}$$

The next step is the permutation of optical paths $2 \to 4$ and $4 \to 2$, which in the second-quantized description takes the form:

$$\sigma_{Bell} : \begin{cases} a_2^\dagger \to a_4^\dagger, \\ a_4^\dagger \to a_2^\dagger. \end{cases} \tag{4}$$

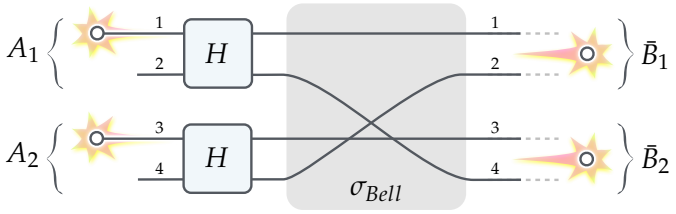


Figure 2: Optical scheme for generation of the Bell state in the no-touching scenario [20] (see also Yurke-Stoler paper [21]).

The protocol uses dual-rail encoding, which means that two output qubits $\bar{B}_1$ and $\bar{B}_2$ are encoded in a pair of paths $\bar{B}_1 = \{1, 2\}$ and $\bar{B}_2 = \{3, 4\}$. For the first qubit $\bar{B}_1$, a particle being present in path 1 corresponds to $a_1^\dagger|0\rangle = |\uparrow_1\rangle$ and a particle being present in path 2 corresponds to $a_2^\dagger|0\rangle = |\downarrow_1\rangle$. Similarly, for the second qubit $\bar{B}_2$. Post-selection procedure consists in checking that there is only a single particle present per each subsystem $\bar{B}_1$ and $\bar{B}_2$ (and rejecting the remaining cases). This guarantees the consistency of dual-rail encoding.

Following all steps of the protocol 1-5, one gets (here, step no. 4 is trivial):

$$a_1^\dagger a_3^\dagger |0\rangle \xrightarrow{H,H} \frac{1}{2}(a_1^\dagger + a_2^\dagger)(a_3^\dagger + b_4^\dagger)|0\rangle \tag{5}$$

$$\xrightarrow{\sigma_{Bell}} \frac{1}{2}(a_1^\dagger + a_4^\dagger)(a_3^\dagger + a_2^\dagger)|0\rangle \tag{6}$$

$$\xrightarrow{post\text{-}sel.} \frac{1}{2}(a_1^\dagger a_3^\dagger + a_4^\dagger a_2^\dagger)|0\rangle\,. \tag{7}$$

where the last line of Eq. (7) corresponds to the state $\frac{1}{2}(|\uparrow\uparrow\rangle \pm |\downarrow\downarrow\rangle)$ (the $\pm$ sign depends on particle's statistics).[1] Note that the sign can be easily locally corrected.

We observe that the obtained state is not normalised due to post-selection in the last step.[2] Therefore, we can read off the efficiency of the protocol by squaring the coefficient in such obtained state, which is equal to Eff = $(1/2)^2 = 1/4$.

2.1.2 Example: Generation of the GHZ state

The GHZ state of three qubits has the following form:

$$|\psi_{GHZ}\rangle = \frac{1}{\sqrt{2}}(|\uparrow\uparrow\uparrow\rangle + |\downarrow\downarrow\downarrow\rangle), \tag{8}$$

where $\{\uparrow, \downarrow\}$ is a computational basis.

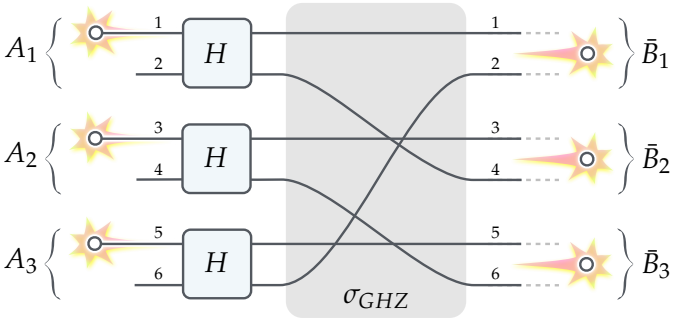


Figure 3: Optical scheme of generating the GHZ state in the no-touching scenario [20] (see also Yurke-Stoler paper [22]).

The optical scheme of generating this entangled state in the no-touching scenario is given in Fig. 3. It starts with the injection of three independent particles into the optical paths 1, 3 and 5, one per each subsystem $A_1 = \{1,2\}$, $A_2 = \{3,4\}$, $A_3 = \{5,6\}$. Then, the particles in each subsystem undergo local unitary transformation by the Hadamard matrix H in Eq. (3). The next step is the permutation of paths σ_{GHZ} given by:

$$\sigma_{GHZ} : \begin{cases} a_2^\dagger \to a_4^\dagger\,, \\ a_4^\dagger \to a_6^\dagger\,, \\ a_6^\dagger \to a_2^\dagger\,. \end{cases} \tag{9}$$

[1] Note that for *bosons* $a_2^\dagger a_1^\dagger = a_1^\dagger a_2^\dagger$ and *fermions* $a_2^\dagger a_1^\dagger = -a_1^\dagger a_2^\dagger$ and in the final sate operators should be in the same order (here we choose the increasing mode labelling order).

[2] Technically, post-selection boils down to rejecting non-compliant terms. In Eq. (6)-(7), the rejected terms are $a_1^\dagger a_2^\dagger$ and $a_4^\dagger a_3^\dagger$ corresponding to two particles present respectively in $\bar{B}_1$ and $\bar{B}_2$.

Here, we use dual rail qubits encoded in pairs of paths $\bar{B}_1 = \{1,2\}$, $\bar{B}_1 = \{3,4\}$ and $\bar{B}_1 = \{5,6\}$. After post-selecting only one particle per each subsystem $\bar{B}_1$, $\bar{B}_2$ and $\bar{B}_3$, the GHZ state of three particles is obtained:

$$a_1^\dagger a_2^\dagger a_3^\dagger |0\rangle \xrightarrow{H,H,H} \frac{1}{2\sqrt{2}}(a_1^\dagger + a_2^\dagger)(a_3^\dagger + a_4^\dagger)(a_5^\dagger + a_6^\dagger)|0\rangle \tag{10}$$

$$\xrightarrow{\sigma_{GHZ}} \frac{1}{2\sqrt{2}}(a_1^\dagger + a_4^\dagger)(a_3^\dagger + a_6^\dagger)(a_5^\dagger + a_2^\dagger)|0\rangle \tag{11}$$

$$\xrightarrow{post\text{-}sel.} \frac{1}{2\sqrt{2}}(a_1^\dagger a_3^\dagger a_5^\dagger + a_4^\dagger a_6^\dagger a_2^\dagger)|0\rangle\,, \tag{12}$$

which is the GHZ state $\frac{1}{2\sqrt{2}}(|\uparrow\uparrow\uparrow\rangle + |\downarrow\downarrow\downarrow\rangle)$ (irrespective of the particle's statistics).

Again, the obtained state is unnormalised due to post-selection. The efficiency of the protocol is equal to $\mathrm{Eff} = (1/2\sqrt{2})^2 = 1/8$.

2.2 Explicit protocols

Let us discuss the generality and practical utility of the *no-touching* scenario, i.e., approach the question:

Q1. *What is the range of entangled states that can be generated in the no-touching scenario? What is the efficiency of those protocols?*

Since the problem of state generation is complex, we took two research directions to illustrate and assess the strength of the *no-touching* proposal. First, we asked about the possibility of generating an arbitrary state of three qubits using the *no-touching* designs [1]. Second, we asked about the performance of the *no-touching* scenarios with the increasing number of qubits on a well-studied example of the W state [2]. In both cases, we gave explicit generation schemes and compared them with other optical proposals in the literature.

2.2.1 Arbitrary state of three qubits

This section presents a briefly overview results from Ref. [1]. The article gives an explicit *no-touching* protocol generating an arbitrary state of three qubits. Our primary motivation behind this paper is getting the broadest possible range of entangled states with minimal resources (just independent particles in the input, passive linear optics and post-selection). To our knowledge, this is the first protocol of this kind in the literature. It should be compared with the protocols based on SLOCCs [19], which start with entanglement in the input (of the GHZ-state and the W-state type) and require complex filtering procedures with efficiency dropping to zero (see Methods in Ref. [1]).

The arbitrary state of three qubits can be written down in the following form:

$$\begin{aligned}|\psi\rangle \;=\; & \gamma_1|\uparrow\uparrow\uparrow\rangle + \gamma_2|\downarrow\uparrow\uparrow\rangle + \gamma_3|\uparrow\downarrow\uparrow\rangle + \gamma_4|\uparrow\uparrow\downarrow\rangle + \gamma_5|\downarrow\downarrow\uparrow\rangle + \gamma_6|\downarrow\uparrow\downarrow\rangle + \\ & \gamma_7|\uparrow\downarrow\downarrow\rangle + \gamma_8|\downarrow\downarrow\downarrow\rangle\,,\end{aligned} \tag{13}$$

where $\{\uparrow, \downarrow\}$ is a computational basis and $\sum_{i=1}^{8} |\gamma_i|^2 = 1$.

We observe that it was shown in Refs. [28, 29] that it is possible to simplify this state using generalised Schmidt decomposition to consist of only five terms. That is, the state in Eq. (13) after some local transformations can be written in the following form:

$$|\psi\rangle \quad = \quad \gamma\,|\uparrow\uparrow\uparrow\rangle + \delta e^{i\varphi}\,|\downarrow\uparrow\uparrow\rangle + \epsilon\,|\downarrow\downarrow\uparrow\rangle + \zeta\,|\downarrow\uparrow\downarrow\rangle + \eta\,|\downarrow\downarrow\downarrow\rangle\,, \tag{14}$$

with five real non-negative parameters $\gamma, \delta, \epsilon, \zeta, \eta$ such that

$$\gamma^2 + \delta^2 + \epsilon^2 + \zeta^2 + \eta^2 = 1\,, \tag{15}$$

and one phase $0 \leq \varphi \leq \pi$. Therefore, it suffices to generate the state in Eq. (14) if one wishes to obtain an arbitrary state of three qubits (it is always possible to use some local transformation on each qubit to transform the obtained state to the form in Eq. (13)).

The optical scheme generating an arbitrary state of three qubits in the form of Eq. (14) is given in Fig. 4. It follows the *no-touching* paradigm from **Section 2.1**:

1. Injection of three independent particles into optical paths 1, 6 and 8 in subsystems A_1, A_2, A_3 respectively.
2. Local unitary transformations U in subsystems A_1 and A_3 and local unitary transformation H in subsystem A_2.
3. Permutation of optical paths $\sigma \in S_{10}$.
4. Local unitary transformation V in subsystem B_1 and local unitary transformation W in subsystem B_2.
5. Post-selection on a single particle in each subsystem $\bar{B}_1$, $\bar{B}_2$ and $\bar{B}_3$.

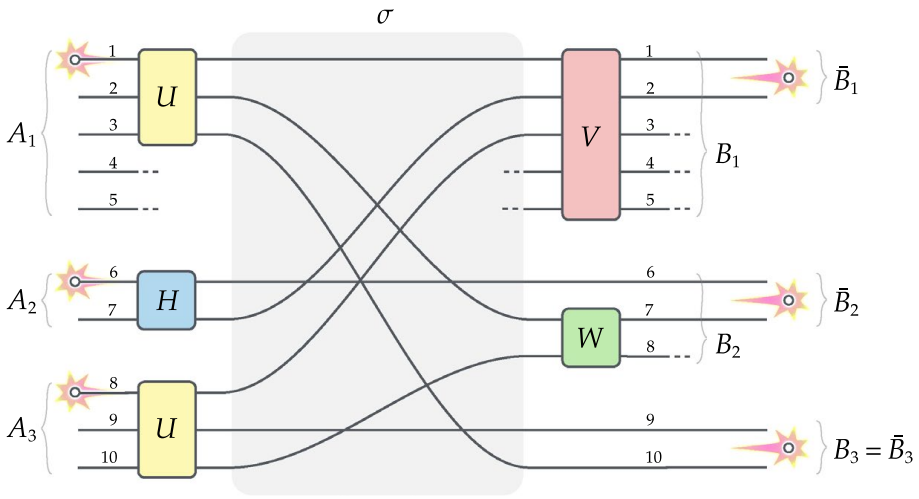


Figure 4: *No-touching* protocol generating an arbitrary state of three qubits [1]. Three independent particles are injected in subsystems A_1, A_2 and A_3 in optical paths 1, 6, and 8. Then they undergo local unitary transformations U, H, U, permutation of optical paths σ and local unitary transformations V and W. The protocol ends with post-selection on the presence of a single particle in each subsystem $\bar{B}_1$, $\bar{B}_2$ and $\bar{B}_3$. An appropriate choice of parameters in unitaries U, H, V and W yields any desired state of three qubits at the output.

After injecting three particles into optical paths 1, 6 and 8, the state undergoes a series of transformations described in Steps 1-4 and post-selected in Step 5. As a result, we get the desired state of three dual-rail qubits encoded in subsystems $\bar{B}_1$, $\bar{B}_2$ and $\bar{B}_3$ is obtained in the form

$$\begin{aligned} a_1^\dagger a_6^\dagger a_8^\dagger \left|0\right\rangle \xrightarrow{1\text{-}5} & \frac{1}{3\sqrt{2}}\left(\gamma a_1^\dagger a_6^\dagger a_9^\dagger + \delta e^{i\varphi} a_2^\dagger a_6^\dagger a_9^\dagger + \epsilon a_7^\dagger b_2^\dagger a_9^\dagger + \zeta a_{10}^\dagger a_6^\dagger a_2^\dagger + \eta a_{10}^\dagger a_2^\dagger a_7^\dagger\right)\left|0\right\rangle \\ & = \frac{1}{3\sqrt{2}}\left(\gamma\left|\uparrow\uparrow\uparrow\right\rangle + \delta e^{i\varphi}\left|\downarrow\uparrow\uparrow\right\rangle + \epsilon\left|\downarrow\downarrow\uparrow\right\rangle + \zeta\left|\downarrow\uparrow\downarrow\right\rangle + \eta\left|\downarrow\downarrow\downarrow\right\rangle\right). \end{aligned} \tag{16}$$

It can be shown that for any given choice of parameters $\gamma, \delta, \epsilon, \zeta, \eta$ and φ, there is an appropriate choice of parameters in the unitaries V and W which gives the desired state Eq. (14) at the output (in this protocol the unitaries U and H are fixed). See Ref. [1] for details. The protocol works for both bosons and fermions (for the latter, the signs can be easily corrected with local unitarties at the output).

One readily reads off the efficiency of the described protocol, which is equal $\mathrm{Eff} = (1/3\sqrt{2})^2 = 1/18 \approx 5\%$. Notably, the efficiency of this scheme is constant for every three qubit-state (it is the same for bosons and fermions). This property gives a considerable qualitative advantage over protocols based on SLOCCs [19] where the efficiency of the filtering procedure drops to zero (see Methods in Ref. [1]). Another remarkable feature is that the presented protocol starts with independent particles in the input. This should be compared with SLOCCs, which require entangled states of the GHZ-state and the W-state type to generate an arbitrary three-qubit state.

In summary, the proposed protocol requires minimal resources (i.e., just three independent particles in the input, passive linear optics and post-selection) and produces an arbitrary three-qubit state with finite and constant efficiency (for bosons and fermions). This seems to be a considerable advancement in state generation techniques (where, for three-qubit entanglement, the dominant approach is based on SLOCCs).

2.2.2 Multipartite W state

This section briefly accounts for the results in Ref. [2]. The article presents an explicit *no-touching* protocol generating the multipartite W state for an arbitrary number of qubits. It is aimed to illustrate how the no-touching designs perform with the increasing number of qubits. We compare our proposal with, to our knowledge, the best existing protocol in Ref. [30], whose efficiency drops polynomially with the increasing number of qubits (cf. schemes in Refs. [31–34] which fall off exponentially).

A multipartite W state of N qubits has the following form:

$$|W_N\rangle = \frac{1}{\sqrt{N}}(|\uparrow\downarrow\downarrow\ldots\downarrow\rangle + |\downarrow\uparrow\downarrow\ldots\downarrow\rangle + \ldots|\downarrow\ldots\downarrow\downarrow\uparrow\rangle), \tag{17}$$

where $\{\uparrow,\downarrow\}$ is a computational basis.

The optical *no-touching* scheme of generation of multipartite W state is presented in Fig. 5. Here are the main steps of the protocol (notice slightly different labeling conventions):

1. Injection of N independent particles into optical paths $1, 2, ..., N$ in each subsystem $A_1, ..., A_N$.
2. Local unitary transformation $U \circ G$ in subsystem A_1 and local unitary transformations V in subsystem $A_2, ..., A_N$.
3. Permutation of optical paths σ.
4. Local unitary transformation G^{-1} on the output of subsystem A_1.
5. Post-selection on a single particle in each subsystem $B_k = \{k, \bar{k}\}$.

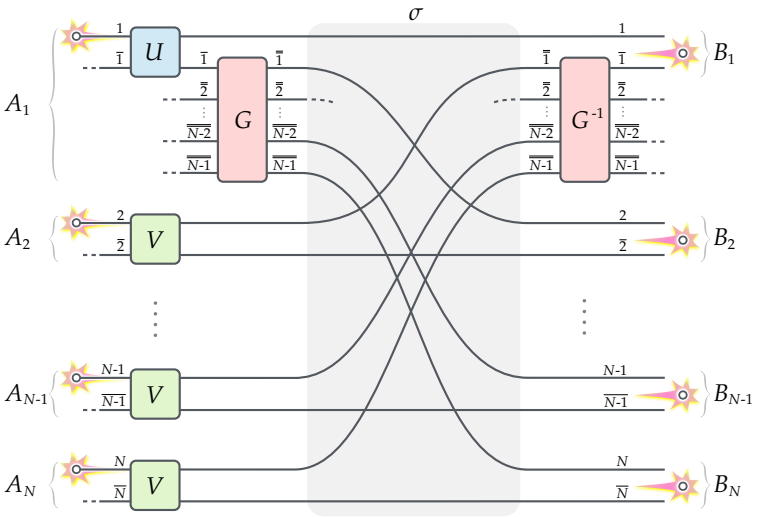


Figure 5: *No-touching* scheme generating multipartite W state. The protocol starts with N independent particles injected into optical paths $1, ..., N$ in the respective subsystems $A_1, ..., A_N$. Then, they undergo local unitary transformations $U \circ G$ and V, permutation of the paths σ and local unitary transformation G^{-1} in subsystem A_1. Post-selection on a single particle in each subsystem $B_1, ..., B_N$ yields the multipartite W in dual-rail encoded qubits.

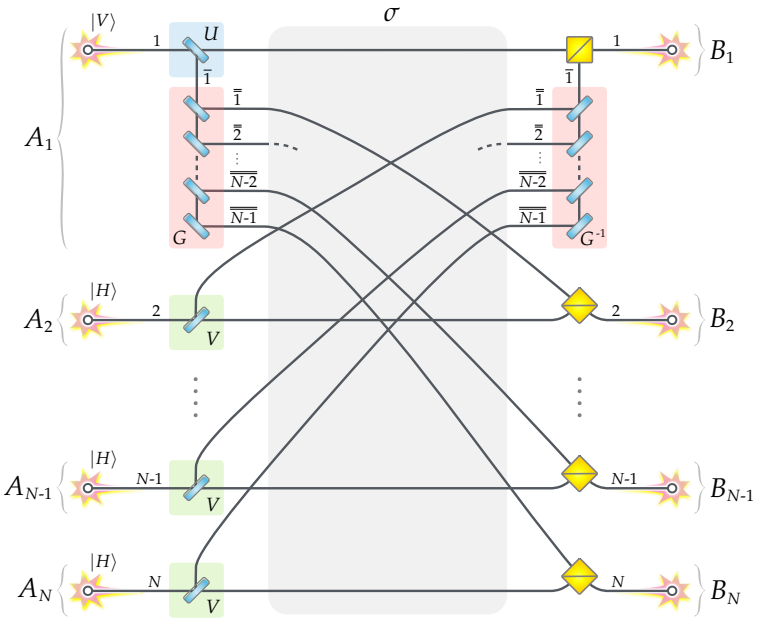


Figure 6: A rewriting of the scheme in Fig. 5 for photon polarisation qubits. Here, the computational basis $\{\uparrow, \downarrow\}$ is encoded in photon polarisation $\{|H\rangle, |V\rangle\}$. The conversion between dual-rail and polarisation encoding is obtained by polarising beam splitters (PBS). As usual, the unitaries are implemented by an appropriate choice of beam splitters. This is analogous of the circuit in Fig. 5 and the analysis is the same.

The protocol starts with injecting N independent particles into the circuit presented in Fig. 5, which implements a series of unitaries in Steps 1-4 and post-selection on a single particle in Step 5. It gives the multipartite W state encoded in dual-rail qubits $B_1, ...B_N$ at the output:

$$
\begin{aligned}
a_1^\dagger a_2^\dagger \dots a_N^\dagger \left|0\right\rangle \quad &\xrightarrow{1\text{-}5} \\
&\sqrt{\frac{\delta^2(1-\delta^2)^{N-1}}{\delta^2+(N-1)^2(1-\delta^2)}}\left(a_1^\dagger a_{\overline{2}}^\dagger \dots a_{\overline{N}}^\dagger + \sum_{i=2}^{k} a_i^\dagger a_{\overline{2}}^\dagger \dots a_{\overline{i-1}}^\dagger a_{\overline{1}}^\dagger a_{\overline{i+1}}^\dagger \dots a_{\overline{N}}^\dagger\right)\left|0\right\rangle \\
&= \sqrt{\frac{\delta^2(1-\delta^2)^{N-1}}{\delta^2+(N-1)^2(1-\delta^2)}}\left(\,\left|\uparrow\downarrow\downarrow \dots \downarrow\right\rangle + \left|\downarrow\uparrow\downarrow \dots \downarrow\right\rangle + \dots + \right. \\
&\qquad\qquad \left. \left|\downarrow \dots \downarrow\uparrow\downarrow\right\rangle + \left|\downarrow \dots \downarrow\downarrow\uparrow\right\rangle\,\right),
\end{aligned}
$$

where δ is a real parameter defined by the appropriate choice of unitaries in the protocol. See Ref. [2] for details. The protocol works for bosons and fermions (in the latter case, the "−" signs due to anti-commutation can be easily corrected by the phase shift in the first path).

Since the obtained state in Eq. (18) is unnormalised due to post-selection, one readily reads off the efficiency of the process as a function of parameter δ and the number of particles N, i.e., we get

$$
\mathsf{Eff}_N(\delta) = \frac{\delta^2(1-\delta^2)^{N-1}}{\delta^2+(N-1)^2(1-\delta^2)}. \tag{18}
$$

Notably, this expression scales polynomially with the increasing number of particles (irrespectively of particles' statistics)

$$
\mathsf{Eff}_N = \max_{\delta} \mathsf{Eff}_N(\delta) \sim \frac{e^{-1}}{N^2} + \frac{7e^{-1}}{2N^3} + o\Big(\frac{1}{N^4}\Big). \tag{19}
$$

There is a qualitative exponential scaling in most of the proposals, see Refs. [31–34]. The only exception is the quantum erasure scheme by Yong-Su Kim et al. [30]. See Fig. 7 for a comparison of the efficiencies of both proposals. In order to give a complete picture we point out that our *no-touching* scheme uses minimal resources, i.e., just passive linear optics, no auxiliary particles, no extra measurements and no active feed-forward state correction. This should be compared with Ref. [30], which uses an auxiliary particle and requires extra measurement to implement quantum erasure. Furthermore, in the most refined version, it resorts to active feed-forward state correction, which counts as a disproportionately large cost in experimental design. See [2] for a discussion.

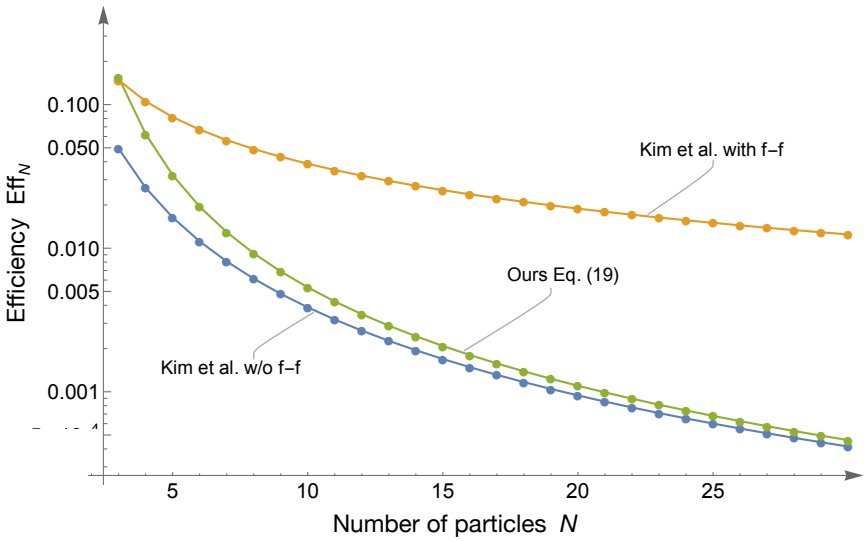


Figure 7: A comparison of the asymptotic behaviour of our scheme with Yong-Su Kim et al. [30]. Our scheme (green line) performs better than the baseline scenario (blue line), requiring an auxiliary particle and extra measurement). We note that with active state correction, the latter scheme improves (orange line). However, this comes with a considerable cost involved with the feed-forward control.

In summary, we gave an explicit *no-touching* protocol generating the W state for an arbitrary number of particles (i.e., just independent particles in the input, passive linear optics and post-selection) whose efficiency decreases polynomially with the number of qubits/particles. The scheme is highly competitive (efficiency and resource-wise) with other proposals in the literature (cf., with exponential scaling of all sequential state-generation techniques [31–34] or other experimental tricks incurring disproportionate costs like feed-forward correction [30]).

3 Some fundamental questions in causal approach

The phenomenon of entanglement has the most profound consequences for our conception of reality, defying some of the most cherished intuitions about how the world works on a fundamental level. It was an ingenious insight of John S. Bell [10, 35, 36], who pointed out that correlations observed in certain experimental setups *cannot* be explained in a causal manner while maintaining, it would seem, natural assumptions. All those predictions were confirmed in a number of sophisticated experiments [37–43].

There are two ways to face the conundrum. In the first reaction, one is inclined to search for the loopholes that might undermine the conclusions drawn from the violation of Bell inequalities [44]. Typically, this boils down to reexamining experimental setups and pointing out discrepancies with the required causal structure. Many of them, like the locality, detection or memory loopholes, were closed in a few recent experiments [37–41]. Some of them, like the free choice or superdeterminism loophole, seem to escape experimental scrutiny (cf. [42, 43] for some interesting proposals). There are also other subtle experimental features which may lead to the so-called *selection bias* [45]. It becomes relevant when the experiment requires *post-selection*, which may introduce non-causal correlations in the collected data. This blurs the validity of the drawn conclusion and requires extra caution in the causal analysis. Notably, the *no-touching* scenarios have post-selection built into the state generation procedure (see Step 5 in **Sections 2.1** and **2.2**). It is thus important to ask the question:

Q2. *Can those protocols be used for conclusive tests of Bell inequalities in the presence of post-selection?*

This problem is extensively discussed in **Section 3.2**. See articles [3, 4].

The second reaction, when it is hard to doubt in the experimental analysis, is to take on board the surprising conclusions from the violation of Bell inequalities. This requires rejecting at least one of the causal assumptions in Bell's theorem, i.e., the *locality* assumption, the *free choice* assumption or the *forward-in-time causation*.[3] See **Section 3.1** for a brief causal account of Bell's theorem. Needless to say, either option (i.e., non-locality [10, 35, 36], lack of free choice/super-determinism [49–51] or retro-causality [52–55]) presents a challenge to our commonsense conception of reality. This is a subject of a serious interpretational debate [11–13], which, to a great extent, draws on one's philosophical preferences. It is thus interesting to take an unbiased position trying to assess the weight of the causal assumptions (hoping to differentiate between them in some quantitative manner). This leads to the last question of a very general nature:

Q3. *What is the causal cost of violation of Bell inequalities, and hence how do the causal assumptions compare one to another?*

[3] For completeness, we mention that one may also take a position dismissing the fundamental role of causality, e.g., treat causality as an emergent concept, try to modify its meaning or the causal framework itself. Here, we take a conservative standpoint where causality is understood in the Pearlian sens [45–48] (essentially the same as introduced by Bell [10]).

The comparison of the locality and free choice assumptions based on the intuitive concept of *fractional measure* is discussed in **Section** 3.3. See article [5].

This section is organized as follows. I begin with a brief causal account of Bell experiment in **Section** 3.1. In **Section** 3.2 the issue of post-selection is discussed. To motivate the problem of *selection bias* in interpreting an experiment, I start in **Section** 3.2.1 with an introductory discussion of the *quantum three-box paradox* [56] with the tools of modern causal framework [45–48]. This serves as an illustration of how elusive the causal conclusions can be when post-selection is an integral part of the experimental design. It is a report on the results in Ref. [3]. Then, in **Section** 3.2.2, I proceed to the analysis of Bell-type experiments with a view to sustaining the conclusions drawn from the violation of Bell inequalities in the presence of certain types of post-selection procedures. This requires reexamining the causal structure in Bell experiments with conditioning on a collider. I report the results in Ref. [4] where a simple criterion, called the *all-but-one principle* was worked out, which safeguards the arguments from Bell inequalities in the post-selected regime. It should be noted that all the protocols presented in **Section** 2 conform to *all-but-one principle*, which safeguards the conclusions about non-locality of the entangled states obtained in these protocols. Finally, in **Section** 3.3, I move to the question about the weight of the assumptions in the Bell theorem. The modern causal framework allows distinguishing those assumptions and comparing them on an equal footing using the concept called *fractional measure* (i.e., the minimal frequency of violation of a given assumption). I explain the results in Ref. [5], where this cost is proved to be equal for the *locality* and the *free choice* assumptions, and explicitly calculated in several relevant cases (including the quantum-mechanical statistics).

3.1 Bell experiment

Let us briefly describe the Bell experiment [10, 35, 36]. It is meant to introduce the basic causal concepts and discuss the assumptions behind Bell inequalities. In the following, we will use the modern causal framework, which has recently gained a solid mathematical foundation in the works of Pearl and others [45–48]. Notably, most of those concepts were implicit in the works of Bell [10], who used the language of hidden (or latent) variables.

Broadly speaking, we are interested in cause-and-effect relationships between the variables, both *observed* or *unobserved* (also called latent or hidden), which are deemed relevant for the description of an experiment. The mathematical framework boils down to the standard probability theory, with the causal relationships being modelled by either functional or (conditional) probability relations between the variables (cf., structural causal models and causal Bayesian models) obeying the so-called Markov condition. Those causal relationships have a graphical representation in terms of the so-called *directed acyclic graphs* (DAGs). See [45–48].[4]

[4]Let me explicitly note that I use the term causal framework interchangeably with the term realist framework. This is congruent with the typical realist position, which adopts the viewpoint

The derivation of Bell inequalities stands on three fundamental assumptions: *locality*, *free choice* and *forward-in-time causation*. Let us be precise about what these assumptions mean. The *locality* assumption postulates that only the immediate surroundings of the laboratory setting should influence the result of the experiment. That is, nothing that is far enough in spatio-temporal sense can have a direct effect on the measurement result (i.e., there are no causal arrows between separated variables). The *free choice* assumption means that the choice of measurement setting can be, in principle, made unaffected by anything relevant to the experiment. This means that the choice variables (also called parameters) cannot be influenced by any other variable (i.e., there are no causal arrows incoming to the variables representing the choice of settings). It is sometimes called *measurement independence* assumption. Finally, there is the *forward-in-time causation* assumption which postulates that all influences should follow the direction of the arrow of time. That is, we do not allow for *retro-causality* (a.k.a. *backwards-in-time causation*). In plain words, *forward-in-time causation* asserts that the causes precede the effects in time.

Having presented all the assumptions, let us consider a typical Bell experiment with two parties, binary settings and binary outcomes. Imagine that the parties, Alice and Bob, are located in their distant (spatio-temporally separated) laboratories. A pair of particles is sent to their laboratories. Each of them makes one of the two possible measurements on their respective particle. Let the variables $x, y = 0, 1$ represent their respective choices (x for Alice and y for Bob). Alice's results are represented by the variable $a = \pm 1$, whereas Bob's results are represented by the variable $b = \pm 1$. Let λ be an *unobserved* (hidden or latent) variable representing any factors from the past which might influence the experimental results. After many repetitions, they meet and compare their results. In this way, a set of probability distributions is constructed $\{P_{ab|xy}\}_{xy}$ describing the probability $P_{ab|xy}$ of obtaining outcomes a and b given the measurement settings x and y. Following the terminology from Ref. [36], we will call the set of probability distributions $\{P_{ab|xy}\}_{xy}$ "a behavior". Note that the collected probability distributions concern only the *observed* variables (i.e., a, b, x, y).

The causal DAG describing the experiment conforming to *locality, free choice and forward-in-time causation* is presented in Fig. 8. If all the assumptions are valid, no additional arrows are allowed. This should be compared with the causal DAGs in Fig. 9, which violate the respective casual assumptions (note that the violation of more than one assumption is also possible).

---

that physical objects and their properties exist independently of the fact of being measured (allowing for those properties to be subsequently disturbed). Notably, it is connected with the notion of counterfactuals which, again, gain a mature account in the modern causal framework [45].

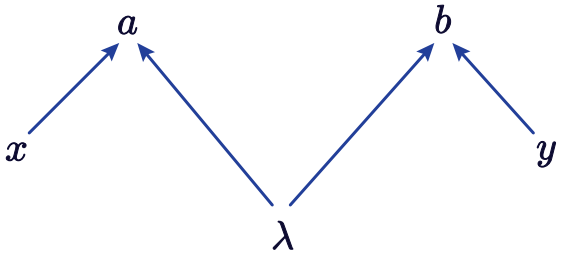


Figure 8: Causal DAG of Bell experiment respecting all three causal assumptions: *locality*, *free choice* and *forward-in-time causation*. No arrows can be added/reversed without violating one of the assumptions.

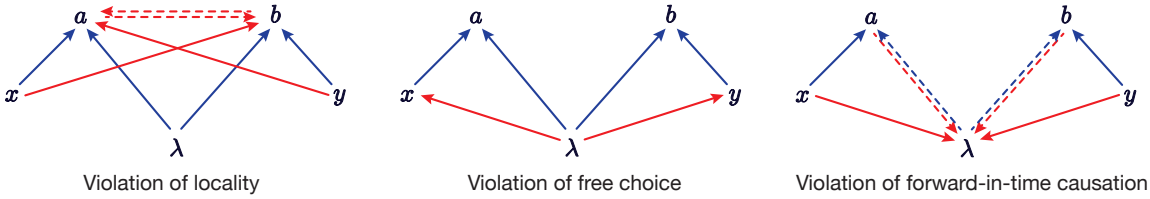


Figure 9: Violation of the respective causal assumptions in Bell experiment. If Bell inequalities are not satisfied, then the causal DAG in Fig. 8 needs to be modified. One red arrow is enough to violate the respective assumption (note that the simultaneous presence of any two parallel dashed arrows is banned by acyclicity of the DAG). Violation of *locality* means an arrow pointing from Alice's side (x,a) to Bob's side (y,b) or vice versa. Violation of *free choice* comes from an arrow pointing to Alice's or Bob's choice of settings variable (x or y). Violation of the *forward-in-time causation* is witnessed by arrows pointing downwards (since, in the picture, the time goes upwards).

The analysis of Bell experiment starts with the usual Bayes' rule, i.e., one can always write

$$P_{ab|xy} = \sum_{\lambda} P_{ab|xy\lambda} \cdot P_{\lambda|xy} \,. \tag{20}$$

Now, suppose that the three causal assumptions hold, that is, the causal structure is described by the DAG in Fig. 8. Then, one immediately reads off the following independence relations. The first is a the factorization called the *locality* condition

$$P_{ab|xy\lambda} = P_{a|x\lambda} \cdot P_{b|y\lambda} \,, \tag{21}$$

which is a consequence of the so-called *outcome independence* and *parameter inde-*

pendence.[5] The second one is the *free choice* condition[6]

$$P_{\lambda|xy} = P_\lambda \qquad \text{(or equivalently } P_{xy|\lambda} = P_{xy}\text{)}. \tag{22}$$

We note that in the causal framework, the conditions Eq. (21) and (22) are a consequence of the postulated DAG in Fig. 8.

Is it possible to obtain the results of the Bell experiment described by behaviour $\{P_{ab|xy}\}_{xy}$, from Eq. (20) with some conditional distributions $P_{ab|xy\lambda}$ and $P_{\lambda|xy}$ satisfying Eq. (21) and (22)? In other words, is the DAG in Fig. 8 compatible with the observed statistics? It turns out that not all distributions can be obtained in such a way, as they have to satisfy the following CHSH inequalities [57]:

$$|S_i| \leq 2, \tag{23}$$

where

$$S_1 = \langle a_0 b_0 \rangle + \langle a_1 b_0 \rangle + \langle a_0 b_1 \rangle - \langle a_1 b_1 \rangle, \tag{24}$$

$$S_2 = \langle a_0 b_0 \rangle + \langle a_1 b_0 \rangle - \langle a_0 b_1 \rangle + \langle a_1 b_1 \rangle, \tag{25}$$

$$S_3 = \langle a_0 b_0 \rangle - \langle a_1 b_0 \rangle + \langle a_0 b_1 \rangle + \langle a_1 b_1 \rangle, \tag{26}$$

$$S_4 = -\langle a_0 b_0 \rangle + \langle a_1 b_0 \rangle + \langle a_0 b_1 \rangle + \langle a_1 b_1 \rangle, \tag{27}$$

where $\langle a_x b_y \rangle = \sum_{a,b} ab P_{ab|xy}$ are the correlations functions. Notably, for binary settings and binary outcomes, the CHSH inequalities are a sufficient condition [58].

Surprisingly, quantum mechanics violates the CHSH inequalities. In general, the CHSH inequalities can be violated up to $|S_i| \leq 4$. In quantum mechanics, they are violated up to the Tsirelson bound [59], i.e. $|S_i| \leq 2\sqrt{2}$, and it is saturated by maximally entangled states (cf. the Bell state Eq. (2)). It means that the experiment cannot be explained via Eqs. (20)-(22). This also indicates that the causal DAG in Fig. 8 is incompatible with the observed statistics, and thus has to be modified if the causal picture is to be maintained.

In the following, in **Section 3.2**, I will discuss certain issues concerning post-selection in the state generation procedure which requires a modification of the causal structure in Fig. 8. In **Section 3.3**, I will quantitatively assess the weight of the violation of the locality and free choice assumptions as depicted in Fig. 9.

[5] *Outcome independence* means $P_{a|bxy\lambda} = P_{a|xy\lambda}$ and $P_{b|axy\lambda} = P_{b|xy\lambda}$, while *parameter independence* means $P_{a|xy\lambda} = P_{a|x\lambda}$ and $P_{b|xy\lambda} = P_{b|y\lambda}$. Both conditions can be read off from the DAG in Fig. 8 as an application of the *d-separation* rules [45], since conditioning on λ blocks all the relevant paths. Then, form the Bayes' rule, we get $P_{ab|xy\lambda} = P_{a|bxy\lambda} \cdot P_{b|xy\lambda} = P_{a|xy\lambda} \cdot P_{b|xy\lambda} = P_{a|x\lambda} \cdot P_{b|y\lambda}$.

[6] Formally, in the language of DAGs it follows from the *d-separation* rules [45], since the variables (x, y) and λ are separated by the colliders a and b.

3.2 Problem with post-selection

Post-selection is commonplace in experimental sciences. It takes place whenever part of the data is rejected in the analysis. While sometimes it may be justified by imperfections of real experiments, it is oftentimes (inconspicuously) built into the experimental design. This is certainly the case in the *no-touching* scenarios in **Section 2**, where post-selection is explicit in Step 5 of the state generation procedure (conditioning on the presence of a single particle in each output channel).

However, post-selection comes with a danger of misinterpreting the data. This is particularly true when reasoning about the cause-and-effect relationships [45]. It is because post-selection can introduce non-causal correlations in the statistics, known as *selection bias* or *Berkson's paradox*, which can distort (or fake) the conclusions drawn from such experiments. It is instructive to give an example.

For illustration, we take a case of Berkson's paradox due to Elwert & Winship [60]. Let us consider two factors, talent (T) and beauty (B) which, in a simplistic picture, are deemed responsible for gaining the status of a celebrity (C). One can observe a correlation between talent and beauty among actors seen at the Oscar gala in Hollywood. If one sees an unattractive actor on the red carpet, it is very likely that he/she is very talented. On the other hand, a very bad and famous actor is likely to be found good-looking. Does it mean that bad looks make one a talented person? Is it a causal mechanism? How does it relate to the fact that in real life (out of Hollywood), we do not observe such a correlation? The explanation comes naturally within the modern language of causal *directed acyclic graphs* (DAGs), where the arrows represent cause-and-effect relationships between the variables. The causal DAG exposing the mechanism behind this situation is shown in Fig. 10. There are three binary variables T (talent), B (beauty) and C (celebrity), with just two arrows $B \to C$ and $T \to C$. There is no direct arrows (causal mechanism) between variables B and T. This follows the intuition that to become a celebrity, one needs to be either very beautiful or very talented (or both), but in general, there is no relation between beauty and talent. The causal DAG explains the lack of correlation between the variables B and T, since the variable C is a collider on the path $B \to C \leftarrow T$ (see the *d-separation* rules in Ref. [45]). However, in the case of conditioning on C, the correlation is likely to appear between the variables B and T (Fig. 10 on the right), since according to the *d-separation* rules conditioning on the collider C removes the block from the path. Thus in the selected population of celebrities, the correlation appears (unless the model is fine-tuned) despite the lack of a causal relationship between talent and beauty. The appearance of *non-causal* correlations in *post-selected* data is known as the *selection bias*.

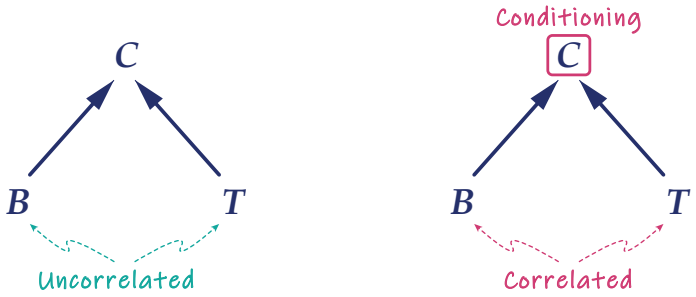


Figure 10: An illustrative example of the *selection bias* (i.e., the appearance of *non-causal* correlations in *post-selected* data). On the left, the variables B (beauty) and T (talent) are uncorrelated since the variable C (celebrity) is a collider which blocks the only path between them. On the right, upon *conditioning/post-selecting* on the variable C (depicted by the red box), the block disappears and the variables B and T become correlated.

The selection bias is an important issue in the analysis of the natural data (when the randomized trials are unavailable), especially when the causal relationships are to be discovered [48]. There is plenty of examples of Berkson's paradox in the research practice, e.g. seemingly beneficial influence of cigarette smoking for dementia [61] or the seemingly beneficial influence of obesity for patients with cardiovascular disease [62]. One should not delude oneself that physics is free of it, as post-selection is oftentimes built into the experimental design.

In the next **Section** 3.2.1, the example of the quantum three box paradox is analyzed to warn against drawing hasty conclusions in the scenarios involving post-selection. We show that careful causal analysis of the paradox may help in explaining the surprising correlations as a case of selection bias. The warning is taken seriously in **Section** 3.2.2 where Bell-type scenarios are reconsidered in the presence of post-selection. In this case, we show that the causal conclusions from the violation of Bell inequalities can be maintained as long as the post-selection obeys the so-called *all-but-one* principle. The latter is a general rule satisfied in a wide range of entanglement generation protocols (including the coincidence-type post-selection with the conserved number of particles in the *no-touching* scenario).

3.2.1 Perplexing post-selection: Quantum three box paradox

This section gives a brief account of the results in Ref. [3] revisiting the quantum three box paradox. It is meant as a warning against jumping to hasty conclusions when interpreting experiments involving post-selection. I explain how to carry out a proper causal analysis of the paradox in the modern language of DAGs [45]. It also reveals interesting features of the causal relationships depending on the preferred viewpoint.

The quantum three box paradox [56] is a prototypical example of the so-called *pre- and post-selection* (PPS) protocols. It was first presented by Aharonov and Vaidman [56] as an illustration of the two-vector formalism and the weak values [63]. It has attracted a lot of discussion and controversy regarding the inter-

pretation and conclusions that can be drawn from this strange effect. See [63–71] for some possible explanations ranging from backwards-in-time influences, non-locality, contextuality, measurement disturbance, etc. All those discussion comes from an apparent contradiction with our commonsense intuitions about the experiment. And it is why it is called a 'paradox'.

However, is it a real paradox which defies any 'reasonable' interpretation? Can't this effect be explained in the usual causal manner? In the following, I will show how the three box paradox can be scrutinised using Pearl's causal framework [45], and argue that the mysterious correlations in this paradox can be explained as a case of a *selection bias*.

(i) Quantum three box paradox

Let us briefly describe the paradox and introduce the notation. See Fig. 11 for illustration of the experiment. It starts with the preparation of the particle in an equal superposition of being present in one of three boxes (*pre-selection*), i.e. the initial state is

$$|\phi\rangle = \frac{|1\rangle + |2\rangle + |3\rangle}{\sqrt{3}}, \tag{28}$$

where the states $|1\rangle, |2\rangle, |3\rangle$ describe the presence of the particle in one of the boxes, $1, 2$ or 3 respectively. At the end of the experiment, one makes the measurement $\mathcal{M}_2$, checking whether that system is in state

$$|\psi\rangle = \frac{|1\rangle + |2\rangle - |3\rangle}{\sqrt{3}}, \tag{29}$$

i.e., the measurement is described by the folloving PVM $\{ |\psi\rangle\langle\psi|, \mathbb{1} - |\psi\rangle\langle\psi| \}$ with the corresponding outcomes $M_2 = 0, 1$. The experiment is considered successful if $M_2 = 1$ (i.e., the system is in state $|\psi\rangle$). If $M_2 = 0$ (i.e., the system is *not* in state $|\psi\rangle$), the experiment is considered a failure, and we reject all those results. This is a *post-selection* procedure.

Now, we ask what happens when we add another *intermediate* measurement $\mathcal{M}_1^C$ checking if the particle is in a given box, where $C = 1, 2, 3$ denotes the *choice* of the opened box. Such a measurement implements the following PVM $\{\mathbb{1} - |C\rangle\langle C|, |C\rangle\langle C| \}$. If the particle is found in a given box C, the outcome is $M_1^C = 1$. Otherwise, if the particle is not found in the box C, the outcome is $M_1^C = 0$.

The surprising results of the quantum three box paradox concern the probabilities of finding a particle in a given box. One may expect that the probability of finding a particle in one of three boxes is equal $1/3$. However, this intuition is wrong. In the experiment described above, the probability of finding a particle in the first box is 100%. One might thus expect that it should be impossible to ever find a particle in one of the other boxes. Here comes a surprise, the probability of finding a particle in box 2 is also 100%.

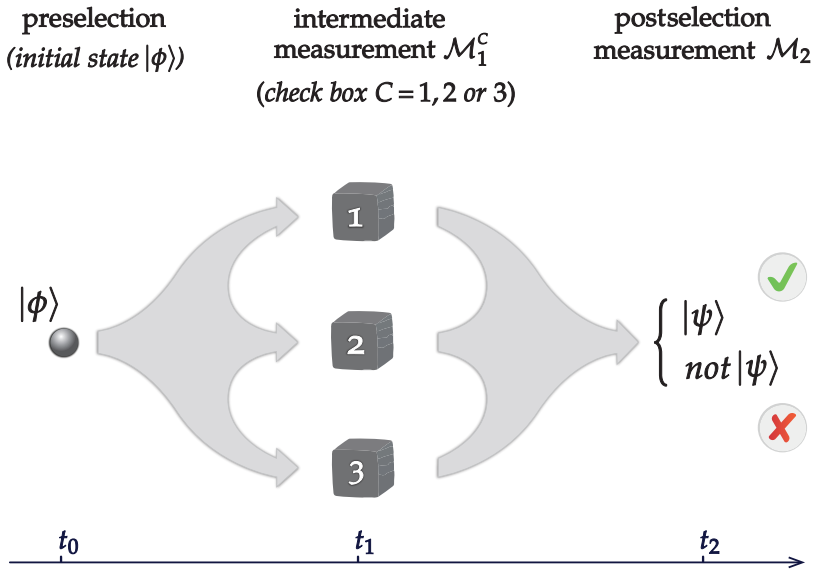


Figure 11: Scheme of the quantum three box experiment. After preparation of initial state $|\phi\rangle$, one chooses to check one of the boxes $C = 1, 2, 3$ in the measurement $\mathcal{M}_1^C$. Then, the final state is *post-selected* on the positive outcome of measurement $\mathcal{M}_2$ checking if the system is in state $|\psi\rangle$. Quantum mechanics gives a surprising result regarding the prediction of finding a particle in boxes 1 and 2: in both cases, the probability of finding a particle is 100% (given successful post-selection).

In order to explicitly calculate the paradoxical probabilities, one uses the so-called Aharonov-Bergmann-Lebovitz (ABL) rule [72]. It indeed gives the probability of finding a particle either in box $C = 1$ or 2 ($M_1 = 1$) equal 100% conditioned on successful post-selection ($M_2 = 1$), i.e.,

$$P(M_1 = 1|M_2 = 1, C) = 1 \quad \text{for} \quad C = 1, 2\,. \tag{30}$$

We observe that in both cases $C = 1$ and 2, the successful post-selection happens with probability $1/9$. For completeness, we give the full experimental statistics in Fig. 12.

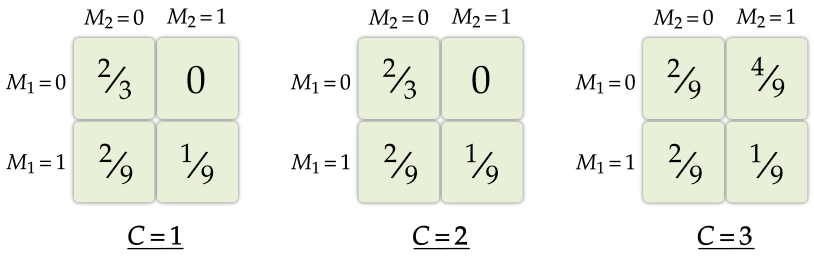


Figure 12: Full statistics in the quantum three box experiment. The tables present all possible joint probability distributions $P(M_1, M_2|C)$ of measurement outcomes ($M_1 = 1$ particle found; $M_1 = 0$ particle not found), and post-selection results ($M_2 = 1$ successful post-selection; $M_2 = 0$ unsuccessful post-selection) for each choice of the opened box $C = 1, 2$ and 3.

We note that in the original formulation of the quantum three box paradox, only the boxes $C = 1$ and 2, i.e. it draws on the correlations in Eq. (30). However, we will see that the full statistics will give more causal insight into the paradox.

(ii) Two interpretational attitudes

There are two positions regarding the causal description one may take.

— Pure causal setting —

This is the most simplistic description of the experiment, i.e., it has the smallest number of variables and assumptions in it. In the *pure causal setting*, we consider only three *observed* variables:

C : choice of the box to be measured ($C = 1, 2, 3$),

M_1: outcome of the box measurement $\mathcal{M}_1^C$ ($M_1 = 0, 1$),

M_2: outcome of the final measurement $\mathcal{M}_2$ ($M_2 = 0, 1$).

In addition, there is one *unobserved* variable:

Λ : hidden (or latent) variable.

The hidden variable Λ is left unspecified so that it can account for any unobserved factors. For instance, it can describe the preparation of the experiment or other factors from the past which may influence the results of experiments.

Note that since the variable Λ is arbitrary, it does not have to be related to the position of the particle before the measurement. In this respect, *pure causal setting* differs from following a more detailed description.

— Realist causal setting —

Here, by *realism* we mean the viewpoint which postulates the existence of the physical objects and their properties independently of the fact of being measured. In this approach, the measurement reveals the preexisting values of the physical property. In the case of the three box experiment, this assumption concerns particle position. This stipulates the existence of another *unobserved* variable:

V : position of the particle ($V = 1, 2, 3$),

which describes the position of the particle *before* the measurement. It is, in principle, different from the hidden variable Λ (which may contain other information).

Given the *realism* assumption, the measurement $\mathcal{M}_1^C$ is supposed to reveal the preexisting property of the particle being in a given box. The following relation should therefore hold true:

$$M_1(C, V) = \delta_{C,V}. \tag{31}$$

where $\delta_{C,V}$ is the Kronecker delta which is equal 1 if the position of the particle V is the same as the the opened box C (otherwise it is equal 0). In this regard, this assumption is more restrictive since it puts additional functional constraints on the variables in the model.

(iii) Causal explanations

The possible causal structures explaining the experiment are presented in Fig. 13. The DAG on the left describes the *pure causal setting*, while the DAG on the right corresponds to the *realist causal setting*. We note that some arrows were excluded based on the two assumptions: only *forward-in-time* causation and *freedom of choice* with regard to measurement settings. The first assumption assures that causation should follow the arrow of time $t_0 < t_1 < t_2$ (i.e., the arrows cannot point backwards in time). The second assumption assures that there are no arrows incoming to the variable C (i.e., the choice cannot be dictated by any other variable).

The blue arrows in Fig. 13 follow the description of the experiment in the *pure* and *realist causal setting*. The red arrows with question marks correspond to the influences that are possibly present in the experiment, i.e. the *outcome dependence* ($M_1 \to M_2$) and *parameter dependence* ($C \to M_2$). Neither of the latter is excluded by any physical principle (like the locality in Bell experiment).

The first observation is that if all the red arrows are present, the full statistics in Fig. 12 can be reproduced by both DAGs in Fig. 13. Hence, the strange behaviour in the quantum three box paradox Eq. (30) can be explained as a case of selection bias. Can we say more about the influences in the models? We follow the principle that the fewer arrows, the better the explanatory power of the DAG. Therefore, it is natural to ask:

> *Which red arrows in Fig. 13 are necessary & enough to explain the statistics observed in the experiment (Fig. 12)?*

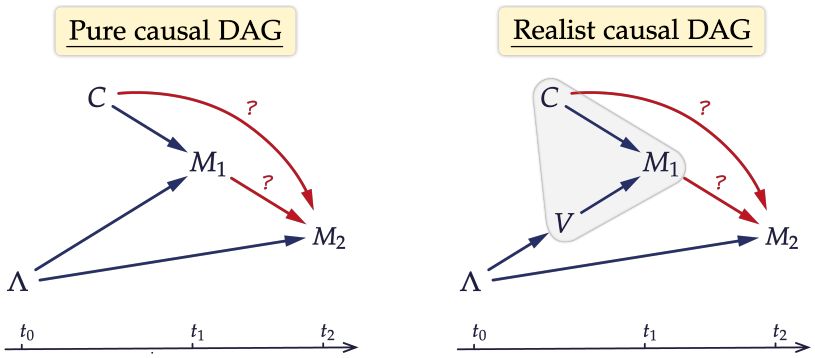


Figure 13: Causal DAGs for the quantum three box paradox. Both diagrams have three observed variables: C - choice of the box measurement, M_1 - result of the box measurement and M_2 - result of the final measurement (post-selection). There are two possible unobserved variables: hidden variable Λ and the position of the particle before measurement V (in the realist approach). The arrows follow forward-in-time causation (there is no retro-causality), and the variable C is free. The grey structure on the right is characteristic of the realist setting, cf. Eq. (31). The red arrows with question marks are possible influences propagating in the experiment: *outcome dependence* ($M_1 \to M_2$) and *parameter dependence* ($C \to M_2$).

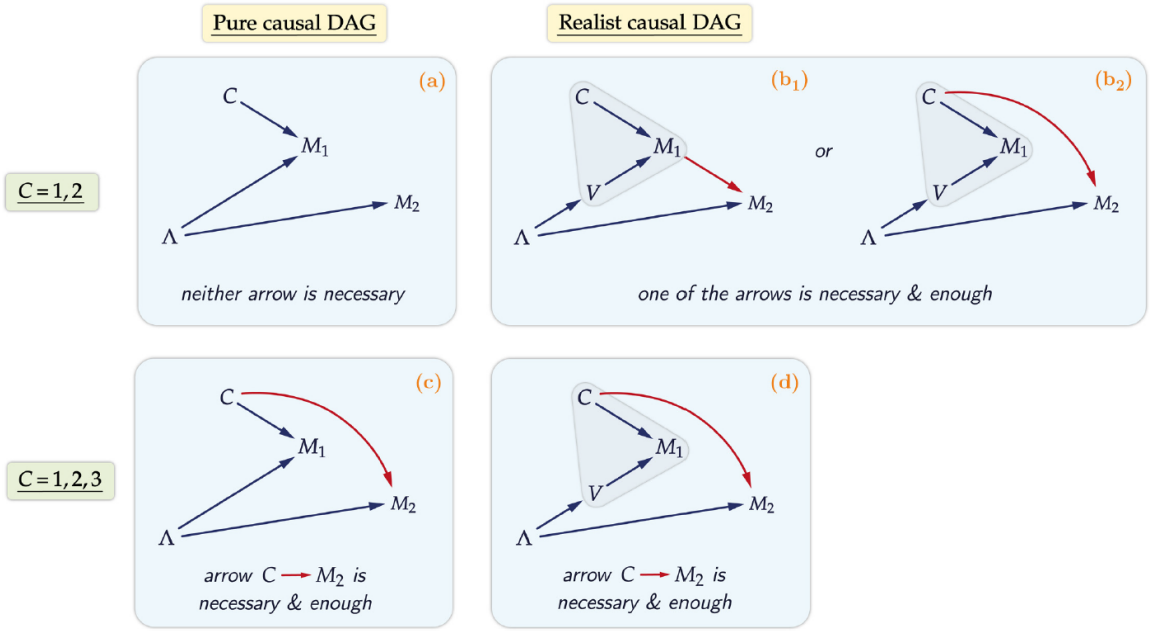


Figure 14: Summary of the results. The causal DAGs represent two approaches: *pure causal setting* and *realist causal setting*. The results differ with regards to chosen statistics, i.e., the full statistics of the experiment $C = 1, 2, 3$ or the statistics restricted to the 'paradoxical' behaviour for $C = 1, 2$. The kind of disturbance needed to explain the experiment, whether it is the *outcome dependence* ($M_1 \to M_2$) or *parameter dependence* ($C \to M_2$), relies on the considered of statistics ($C = 1, 2, 3$ or $C = 1, 2$) and the preferred worldview (*pure* or *realist causal setting*).

The answer is succintly summarised in Fig. 14. The presented DAGs display the red arrows proved to be necessary and enough to explain the experiment. In each case, the sufficiency can be justified by the explicit construction of the corresponding *structural causal model* (SCM). The necessity follows from the analysis of *instrumental inequalities* [45, 73]. See Ref. [3] for a discussion and details. Note that the results differ depending on one's preferred approach, i.e. *pure* vs. *realist causal setting*. Interestingly, the full statistics give more insight into the structure of influences in the experiment (cf., the necessity of *parameter dependence* vs the explanation *without* any extra dependencies for the restricted 'paradoxical' statistics in the pure causal setting). The presence of direct influence on the result of post-selection in each causal DAG (whether it is caused by hidden variable lambda, *parameter dependence* or *outcome dependence*), indicates that we can indeed deal here with the instance of Berkson's paradox.

3.2.2 All-but-one principle as a warrant of safe post-selection

This section is a brief report on the results in Ref. [4] concerning the arguments from Bell inequalities in the presence of post-selection.

As illustrated above in the above examples, post-selection may be responsible for additional correlations and even lead to paradoxical interpretations (which, in

fact, can be explained as a consequence of processing the data and not an effect of a real causal nature). This phenomenon is known in the literature as *selection bias* or *Berson's paradox* [45]. The problem also arises in Bell-type arguments in experiments using post-selection, which can raise doubts if the obtained results support the causal conclusions, or are just an effect of processing the data. This section will present a simple criterion called the *all-but-one principle* which shows when one can say that post-selection in Bell-type test is *safe*, i.e., when the causal conclusions from the violation of Bell inequalities remain in force.

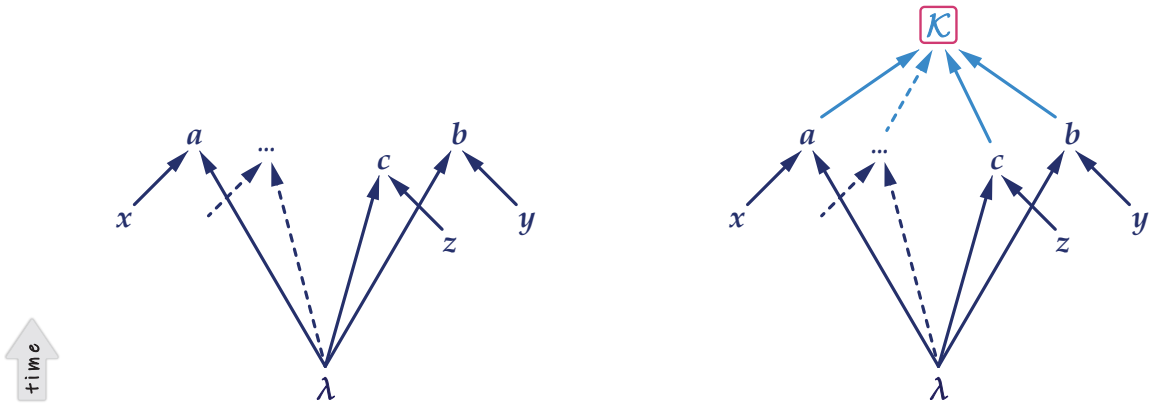


Figure 15: Causal structure of multipartite Bell experiment *without* post-selection procedure (on the left) and *with* post-selection procedure (on the right). Measurement outcomes $a, b, c, ...$ are influenced by measurement settings $x, y, z, ...$ as well as hidden variable λ. In the causal DAG on the right, there is an additional variable $\mathcal{K}$ (depending on $a, b, c, ...$) which decides about post-selection (*conditioning* on $\mathcal{K}$ determines the post-selected ensemble). Since $\mathcal{K}$ is a collider, it may introduce additional (non-causal) correlations in the post-selected data.

The causal DAG describing the multipartite Bell experiment *without* post-selection is shown in Fig. 15 (on the left). In an analogous way to the bipartite scenario, we consider the experiment with many parties $A, B, C, ...$ with the respective measurement settings $x, y, z, ...$ and outcomes $a, b, c, ...$. The observed probability distributions will be noted as $\{P_{abc...|xyz...}\}_{xyz...}$ which is a straightforward generalisation from the bipartite case presented in **Section 3.1**.

A general form of Bell inequalities in the multipartite scenario is as follows

$$\sum_{\substack{abc...\\xyz...}} s^{abc...}_{xyz...} P_{abc...|xyz...} \leq \mathcal{I}_L \tag{32}$$

where $s^{abc...}_{xyz...}$ and $\mathcal{I}_L$ are some numbers. Cf. Eq. (23). They are proved in the usual manner by assuming the existence of some hidden variable λ, using the Bayes rule (cf. Eq. (20))

$$P_{abc...|xyz...} = \sum_{\lambda} P_{abc...|xyz...\lambda} \cdot P_{\lambda|xyz...}, \tag{33}$$

and postulating the *locality* and *free choice* assumptions. The *locality* assumption boils down to the factorisation (cf. Eq. (21))

$$P_{abc...|xyz...\lambda} = P_{a|x\lambda} \cdot P_{b|y\lambda} \cdot P_{c|z\lambda} \cdot ..., \tag{34}$$

which asserts that each measurement outcome $a, b, c, ...$ is influenced (depend on) only by its respective setting $x, y, z, ...$ and the hidden variable λ. The *freedom of choice* assumption (sometimes called *measurement independence*), states that the measurement settings and hidden variable λ should not influence (depend on) one another, i.e. we have

$$P_{\lambda|xyz...} = P_\lambda \qquad \text{(or equivalently } P_{xy...|\lambda} = P_{xy...}\text{)}. \tag{35}$$

It is straightforward to see that Eqs. (34) and (35) follow from the causal DAG in Fig. 15 on the left (cf. *d-separation* rules).

Now, let us consider a variant of the multipartite Bell experiment *with* post-selection. In general, the structure of outcomes $a, b, c, ...$ is arbitrary (in particular, the outcomes may include non-results, like no-click in a detector). The essence of post-selection is to neglect certain results in the analysis of the experiment (like the lack of click or coincidence in the detectors). Sometimes it is justifiable (e.g., accidental loss or imperfections of the experiment). But, sometimes, it is deliberately built into the very bones of the experimental design (e.g., the requirement of the coincidence clicks in the detectors). Post-selection boils down to collecting all the experimental data and rejecting those experimental trials which do not meet a given criterion. Formally, this procedure can be described as conditioning on some binary variable

$$\mathcal{K} \equiv \mathcal{K}(a, b, c, \dots), \tag{36}$$

which is a function of all possible outcomes $a, b, c, ...$. We use the convention where $\mathcal{K} = 1$ corresponds to the successful post-selection, i.e. the result is kept, and $\mathcal{K} = 0$ corresponds to the unsuccessful post-selection, i.e. the result is rejected. Note that in the case of Bell experiment, the result of post-selection $\mathcal{K}$ is decided only after all the parties meet with each other, which corresponds to the structure of the causal DAG in Fig. 15 on the right (and hence all the arrows from $a, b, c, ...$ collide in $\mathcal{K}$).

From the causal point of view, post-selection may cast some doubts about the validity of conclusions drawn from the violation of Bell inequalities. Indeed, Bell inequalities are derived from the DAG on the left in Fig. 15, while the actual DAG (on the right) involves conditioning on a collider $\mathcal{K}$. We know that this may result in a *selection bias* introducing additional non-causal correlations in the data [45], and thus fake the conclusions from the Bell test. Therefore, we introduce the concept of *safe post-selection* which *'does not undermine the conclusions from Bell inequalities Eq. (32) calculated from the post-selected data'*. To be more precise, since the derivation of Bell inequalities draws on the two assumptions Eq. (34) and (35), we make the following formal definition.

Definition 1 *Post-selection procedure described by the variable $\mathcal{K}(a, b, c, \ldots)$ is considered to be safe if the locality assumption and the free choice assumption hold in the post-selected regime ($\mathcal{K} = 1$), i.e.,*

$$P_{abc\ldots|xyz\ldots\mathcal{K}} = P_{a|x\lambda\mathcal{K}} \cdot P_{b|y\lambda\mathcal{K}} \cdot P_{c|z\lambda\mathcal{K}} \cdot \ldots , \tag{37}$$

and

$$P_{\lambda|xyz\ldots\mathcal{K}} = P_{\lambda|\mathcal{K}} . \tag{38}$$

The natural problem to pose is whether we can *give a principled answer to the question when post-selection may be considered safe*. In order to answer that question, we introduce the following criterion called the *all-but-one* principle.

Definition 2 *Post-selection defined by the function $\mathcal{K}(a_1, ..., a_N)$ follows the all-but-one principle if its result can be fully determined without knowing one of the outcomes, i.e.,*

$$\mathcal{K} = \mathcal{K}(a_1, ..., \not{a}_k, ..., a_N) , \tag{39}$$

for each $k = 1, ..., N$. The slash over $\not{a}_k$ stands for the missing outcome of k-th party.

Informally, one can say that in the case of Bell scenario, it is enough to know the outcome of all parties except one to determine if post-selection was successful ($\mathcal{K} = 1$) or not ($\mathcal{K} = 0$). For instance, in the case of three party Bell scenario, the result of post-selection can be determined knowing only two outcomes:

$$\mathcal{K}(a, b, c) = \mathcal{K}(a, b) = \mathcal{K}(a, c) = \mathcal{K}(b, c). \tag{40}$$

Now, we are ready to state the main result of this section that is proved Ref. [4].

Theorem 1 *For an arbitrary number of parties, post-selection which follows all-but-one principle is always safe.*

This criterion safeguards conclusions from the violation of Bell inequalities in the post-selected regime. It is particularly helpful in the protocols with the conserved number of particles and coincidence-type post-selection, which are often used in quantum optics, cf. [14, 15]. Examples of such protocols include, for instance, entanglement by path identity [74, 75], spatial overlap of indistinguishable particles [25, 76] or integrated photonic techniques [16, 77]. We note that the *all-but-one* principle concerns post-selection built into the design (e.g., it does not apply to the detection loophole, which does not obey it).

Most importantly, all the *no-touching* scenarios presented in **Section 2** have post-selection in the design (see Step 5), which conforms to *all-but-one* principle (i.e., coincidence on a single particle in each channel with the conserved number of particles). This means the states generated in such protocols can be used for Bell non-locality tests without casting doubts as regards the *selection bias*.

3.3 Causal cost of violating Bell inequalities

In the last part, let us discuss and compare the weight of Bell assumptions, presenting a short overview of the main ideas and results in Ref. [5].

Violation of Bell inequalities shows that at least one of the assumptions of the experiment (i.e., *locality*, *free choice* or *forward-in-time causation*) must be discarded if the standard causal picture is to be maintained. See a discussion in **Section 3.1**. Here, instead of debating the philosophical grounds for rejecting a given assumption [11–13], we will focus on a more modest mathematical question

Q3. *What is the causal cost of violation of Bell inequalities, and hence how do the causal assumptions compare one to another?*

More specifically, we will focus on two assumptions comparing the cost of violating *locality* vs *free choice* (i.e., we do not consider retro-causal models). In the following, we use the standard causal/hidden variable framework as discussed in **Section 3.1**. Such an approach allows for a clear distinction between the causal assumptions which comes down to the presence of certain arrows in the DAG, cf. Fig. 9. Such an approach allows for a quantitative comparison of the violation of assumptions based on the following research question:

> *How often one needs to relax a given assumption while keeping the rest intact (by adding an appropriate arrow), in order to explain the observed statistics?*

Let us consider a standard bipartite Bell scenario with Alice and Bob having a multiple choice of settings $x, y \in \mathfrak{M} = \{1, 2, \ldots, M\}$, and their respective outcomes are labelled as $a, b = \pm 1$. See **Section 3.1**. When Alice and Bob compare their results, the outcomes are collected in the set of $M \times M$ probability distributions $\{P_{ab|xy}\}_{xy}$ called a "behaviour", where $P_{ab|xy}$ stands for the conditional probability of obtaining the results a and b given the choice of settings x and y. They can also calculate their distribution of settings P_{xy}. Therefore, to explain the statistics observed in a given experiment, one needs to account for both

$$\begin{aligned} \{P_{ab|xy}\}_{xy} &\quad - \quad \text{the "behaviour", and} \\ P_{xy} &\quad - \quad \text{the distribution of settings.} \end{aligned}$$

It is important to note that P_{xy} can in principle be chosen arbitrarily.

In the following, we will be interested in the smallest frequency of the violation of a given assumption that is needed to reproduce the experimental statistics. We are guided by the principle that the less violation, the better. So, we define two natural measures based on the concept of simulability in a repeated experiment. The first one is the measure of locality μ_L (also called a *local fraction*)

μ_L – *the maximal fraction of trails without violation of locality which can reproduce a given behaviour* $\{P_{ab|xy}\}_{xy}$ *for arbitrary distribution* P_{xy}*, optimised over all possible strategies with freely chosen settings.*

The second one is the measure free choice μ_F (also called a *free fraction*), which is defined in a similar manner

μ_F – *the maximal fraction of trails with freely chosen experimental settings which can reproduce a given behaviour* $\{P_{ab|xy}\}_{xy}$ *for arbitrary distribution* P_{xy}*, optimised over all possible local strategies.*

Such a defined measure of locality was proposed by Elitzur, Popescu and Rohrlich in Ref. [78] and developed in Refs. [79–83]. As for the measure of free choice, it was the first time discussed in Ref. [5]. All other approaches to measuring free choice in the literature are essentially different, cf. Refs. [84–90]. We note that, for the purpose of comparison, it is crucial to have both measures defined on the same conceptual basis (here, it is the notion of fractional measure).

Below, we briefly discuss the distinction between *local hidden variable* (LHV) model and the *free hidden variable* (FHV) models on which the measures μ_L and μ_F are based and connect them with the causal concepts.

3.3.1 Two types of hidden variable models

Following the discussion in **Section 3.1**, we will consider four observed variables x, y (choice of settings), a, b (measurement outcomes) and some unobserved hidden variable λ (assumed to describe any factors in the past). Defining a hidden variable model boils down to the specification of the hidden variable space $\Lambda(\ni \lambda)$ and postulating conditional dependencies between the variables which follow the stipulated causal structure. Here, we are interested in models which obey forward-in-time causation and allow a violation of just one of the remaining assumptions, i.e., either locality or free choice. See Fig. 9 for illustration.

— Free hidden variable model (FHV) —

The FHV model preserves the *free choice* assumption at each experimental trial. This assumption (sometimes called *measurement independence*) is typically stated as the following independence condition

$$P_{\lambda|xy} = P_\lambda \qquad \text{(or equivalently } P_{xy|\lambda} = P_{xy}\text{)}, \tag{41}$$

for all measurement settings $x, y \in \mathfrak{M}$ and all $\lambda \in \Lambda$. It means that the hidden variable λ should not contain any information about Alice and Bob's settings.

In the causal language, the above condition follows from the requirement that Alice and Bob's choices are not affected by any other variable (no causal arrows incoming to x and y). Let us enumerate all possible scenarios where *free choice* assumption is preserved. The first one is the trivial Bell structure in Fig. 8 (where *locality* is also preserved). All DAGs with a single extra arrow are enumerated in Fig. 16. One may go further by combining the red arrows into more complex DAGs as long as acyclicity is preserved, cf. Fig. 9 (on the left). We call all those structures *free hidden variable* (FHV) models. Clearly, in such models Eq. (41) is satisfied.

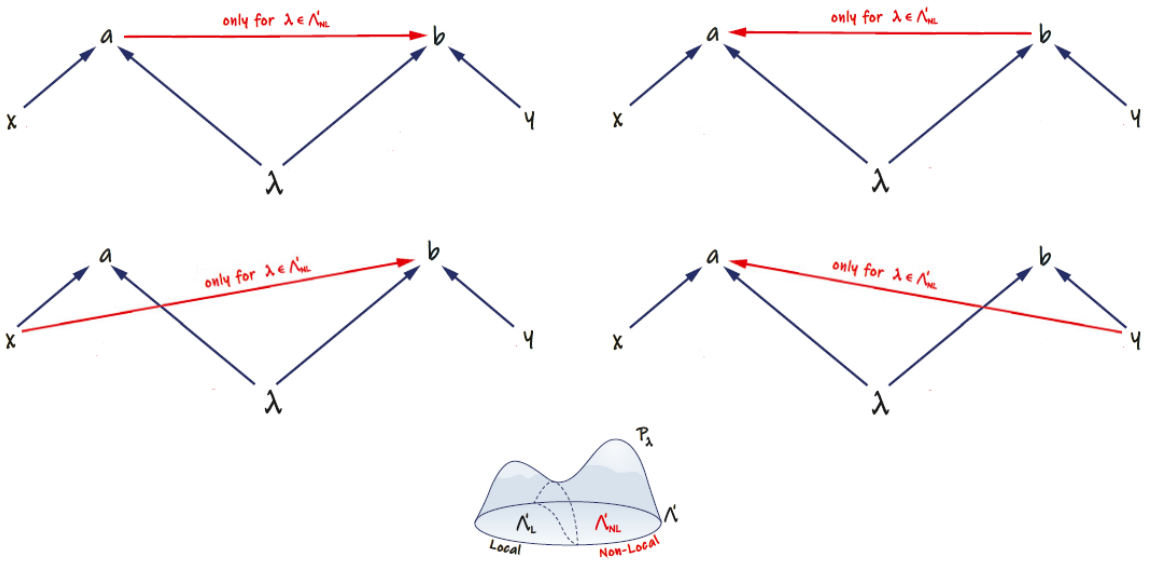


Figure 16: Four possible causal DAGs for Bell experiment with *freedom of choice* and a single *non-local* arrow (depicted in red). For a given *free hidden variable* (FHV) model we can differentiate between the λs for which at least one red arrow is present ($\lambda \in \Lambda'_{NL}$) and those for which all red arrows are missing ($\lambda \in \Lambda'_{L}$). The cumulative distribution $\sum_{\lambda \in \Lambda'_L} P_\lambda$ defines the measure of locality in the model.

Note that in those graphs each red arrow $a \to b$, $a \leftarrow b$, $x \to b$ or $y \to a$ violates the *locality* assumption (i.e., some variable on Alice's side affects some a variable on Bob's side, or vice versa). It is crucial to observe that those non-local arrows may be effective only *sometimes*. This means that, in the repeated experiment, for some λs is some red arrow present while for others all red arrows are absent. This leads to the splitting of the hidden variable space $\Lambda' = \Lambda'_L \cup \Lambda'_{NL}$.[7] The subspace Λ'_L describes the cases when *locality* holds in the FHV model. The subspace Λ'_{NL} describes the cases when *locality* is violated in the FHV model. Formally, those subsets are defined as follows:

$$\lambda \in \Lambda'_L \iff \text{Eq. (21) holds for all } x, y \in \mathfrak{M}, \tag{42}$$

$$\lambda \in \Lambda'_{NL} \iff \text{Eq. (21) fails for some } x, y, \in \mathfrak{M}. \tag{43}$$

— Local hidden variable model (LHV) —

The LHV model preserves *locality* assumption at each experimental trial. The *locality* assumption is usually defined as the following factorization condition

$$P_{ab|xy\lambda} = P_{a|x\lambda} \cdot P_{b|y\lambda}, \tag{44}$$

[7] To avoid ambiguity, we denote the hidden variable space Λ for FHV models with a prime Λ', and for LHV models with a double prime Λ''. This avoids the confusion between the subsections treating different types of models.

for all measurement settings $x, y \in \mathfrak{M}$ and all $\lambda \in \Lambda$. It means that the full information about the past λ screens off Alice's and Bob's sides of the experiment.

In the causal framework, this condition follows from the lack of arrows from Alice's side (variables x and a) to Bob's side (variables y and b) and vice versa. It is justified by the spatio-temporal separation of the respective regions. Clearly, the trivial Bell structure in Fig. 8 falls into that category (in that case *free of choice* is also preserved). There are also two possibilities with a single arrow depicted in Fig. 17 and one DAG with both red arrows together, cf. Fig. 9 (in the middle). Those structures are called *local hidden variable* (LHV) models. Such models satisfy Eq. (44).

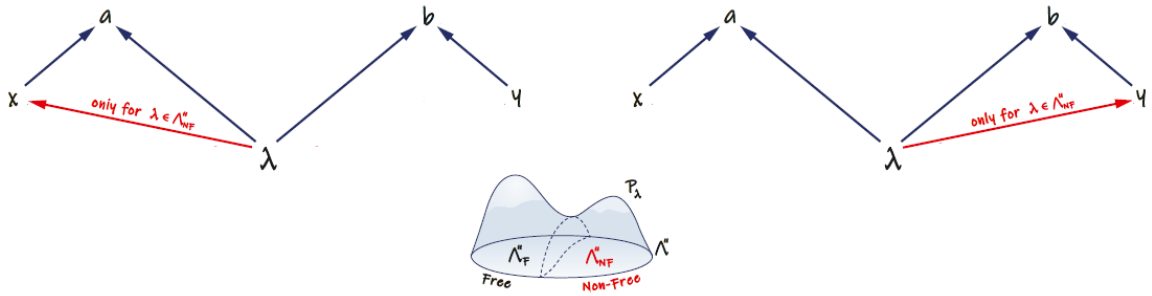


Figure 17: Two possible causal DAGs for Bell experiment preserving *locality* and a single arrow violating *freedom of choice* (depicted in red). For a given *local hidden variable* (LHV) model we can distinguish between the λs for which at least one red arrow is present ($\lambda \in \Lambda''_{NF}$) and those for which all red arrows are absent ($\lambda \in \Lambda''_{F}$). Calculating the cumulative distribution $\sum_{\lambda \in \Lambda''_F} P_\lambda$ gives the measure of free choice in the model.

It is straightforward to see that each red arrow $\lambda \to x$ and $\lambda \to y$ violates the *free choice* assumption (i.e., the past affects Alice and Bob's choices). Again, we observe that this influence may happen only *sometimes*. That is, in the repeated experiment, for some λs is some red arrow present while for others all red arrows are gone. Therefore, the hidden variable space splits $\Lambda'' = \Lambda''_F \cup \Lambda''_{NF}$.[7] The subspace Λ''_F describes the cases with *freedom of choice* in the LHV model. The subspace Λ''_{NF} describes the cases when *free choice* is violated in the LHV model. This is formally defined as follows:

$$\lambda \in \Lambda''_F \iff \text{Eq. (22) holds for all } x, y \in \mathfrak{M}, \tag{45}$$

$$\lambda \in \Lambda''_{NF} \iff \text{Eq. (22) fails for some } x, y, \in \mathfrak{M}. \tag{46}$$

3.3.2 Measures of locality and free choice

Having introduced two classes of models LVH and FHV, we can define measures of locality μ_L and free choice μ_F.

Definition 3 *For a given behavior $\{P_{ab|xy}\}_{xy}$ the measure of locality μ_L and the measure of free choice μ_F are defined in the following way:*

$$\mu_L := \min_{P_{xy}} \max_{FHV} \sum_{\lambda \in \Lambda_L} P_\lambda \,, \tag{47}$$

$$\mu_F := \min_{P_{xy}} \max_{LHV} \sum_{\lambda \in \Lambda_F} P_\lambda \,, \tag{48}$$

where the maxima are taken over all the respective LHV and FHV models simulating the behaviour $\{P_{ab|xy}\}_{xy}$ with a fixed distribution of settings P_{xy}, and the minima are taken over all possible choices of the latter.

The maxima over FHV/LHV account for the best possible model which gives the highest fraction of trials with *locality/free choice* retained (where in the whole experiment the respective *free choice/locality* is fully maintained). The minimum over P_{xy} accounts for the worst-case scenario (since the choice of settings can be in principle arbitrary).

It turns out that the minima in Eq. (47) and Eq. (48) can be omitted (see Ref. [5] for the proof). The **Definition 3** can be therefore simplified in the following way:

Lemma 1 *The minima in **Definition 3** in Eq. (47) and Eq. (48) can be omitted. Hence, the measures of locality μ_L and free choice μ_F have the following form*

$$\mu_L = \max_{FHV} \sum_{\lambda \in \Lambda_L} , P_\lambda \tag{49}$$

$$\mu_F = \max_{LHV} \sum_{\lambda \in \Lambda_F} , P_\lambda \tag{50}$$

where the maxima are taken for some fixed nontrivial distribution of settings P_{xy}.

Now, we are ready to present the first result in Ref. [5] concerning the comparison of weights of the respective assumptions. We pose the question about the frequency of violations of *locality* and *free choice* in the Bell test. *How often one needs to relax one of the assumptions in order to recover the experimental statistics?* It is surprising to see that, in fact, the frequency of violation of either free choice assumption or locality assumption is the same.

Theorem 2 *For a given behaviour $\{P_{ab|xy}\}_{xy}$, the degree of locality and freedom of choice are the same, i.e. the measures in **Definition 3** are equivalent $\mu_F = \mu_L$.*

This result holds for an arbitrary number of settings $x, y = 1, 2, 3, \ldots$. It also extends to a larger number of outcomes $a, b = 1, 2, 3, \ldots$, and many parties Alice, Bob, Charlie etc. See Ref. [5] for the proof.

From **Theorem 2** it follows that the violation of either locality or free choice is equally costly. At least from the perspective of causal graphs and frequency measures, they should be treated on an equal footing. It means that the preference for one assumption over the other should come from the details of the experimental protocol or one's ontological viewpoint.

This result has two important features. Firstly, it is *theory-independent* in the sense that it refers only to the experimental statistics without bothering with the details of the experimental protocol or an underlying theoretical model (with the quantum mechanical predictions being just one possible field of applications). Secondly, the *interchangeability* of both measures μ_L and μ_F have a practical advantage: knowledge of one allows to immediately infer the value of the other.

We note that **Theorem 2** can be seen as a general structural property of causal modelling of Bell-type experiments. One can inquire further asking if those measures can be calculated explicitly. Here is the second general result in Ref. [5].

Theorem 3 *For a given non-signaling*[8] *behavior* $\{P_{ab|xy}\}_{xy}$ *in the bipartite Bell scenario with binary settings* $x, y \in \mathfrak{M} = \{0, 1\}$, *both measures* μ_L *and* μ_F *in* ***Definition 3*** *are equal to:*

$$\mu_L = \mu_F = \begin{cases} \frac{1}{4}(4 - S_{max}), & \text{if } S_{max} > 2, \\ 1, & \text{otherwise.} \end{cases} \tag{51}$$

where $S_{max} = max\{|S_i| : i = 1, \ldots, 4\}$ *is the maximal absolute value of the four CHSH expressions in Eq. (24)-(27).*

Notably, this is an *explicit* and *theory-independent* method of calculating such defined degree of *locality* and *free choice*. It refers solely to the experimental statistics without requiring the specifics of the experimental design or the underlying theory. Furthermore, **Theorem 3** provides a simple interpretation of the amount of violation of the CHSH inequalities in Bell experiment as a fraction of trials when either *locality* or *free choice* needs to be violated to explain the observed results. We note that so far the violation of the CHSH inequalities has been treated as a binary test, while now we give a precise meaning to the amount of violation.

3.3.3 The quantum-mechanical statistics

So far we have presented the model-independent results regarding Bell-type experiments. In the following, we will show what happens when one restricts oneself to the case of quantum-mechanical statistics.

[8] Non-signalling means that Bob's choice (whether y=0 or y=1) does not influence the statistics on Alice's side, i.e. $P_{a|x0} = \sum_b P_{ab|x0} = \sum_b P_{ab|x1} = P_{a|x1}$ for all x, a. The same applies to the other direction where the statistics on Bob's side is not affected by Alice's choice (whether x=0 or x=1), i.e. $P_{b|0y} = \sum_a P_{ab|0y} = \sum_a P_{ab|1y} = P_{b|1y}$ for all y, b.

First, let us consider maximally entangled states of two qubits. **Theorem 3** gives us a way to calculate the maximal violation of *locality* or *free choice* for bipartite Bell test with binary settings $x, y \in \mathcal{M} = \{0, 1\}$. We can immediately use Tsirelson's bound [59] $S_{max} = 2\sqrt{2}$ and plug it into Eq. (51) to see that both fractional measures are equal $\mu_L = \mu_F = 2 - \sqrt{2} \approx 59\%$ for maximally entangled states of two qubits (and appropriate measurements). For other states and measurements, it can be also easily calculated.

Another interesting result concerns the case when we increase the number of experimental settings up to infinity, i.e. when Alice's and Bob's settings are $x, y \in \mathcal{M} = \{1, 2, 3, \ldots, M\}$ with $M \to +\infty$. It was shown in Refs. [78, 80] that the *local fraction* drops to zero with the increasing number of experimental settings. This immediately allows a translation to the case of the measure of *free choice*.

Theorem 4 *The free fraction μ_F in every local hidden variable model explaining Bell experiment with maximally entangled states, tends to zero as the number of experimental settings increases $M \to +\infty$, i.e. $\mu_L = \mu_F \xrightarrow[M \to +\infty]{} 0$.*

Finally, let us consider less entangled states with some arbitrary number of settings. For a pure not maximally entangled two-qubit state, the amount of *free choice* is the same as *local fraction* which is known for all pure two-qubit states, see Ref. [83]. Thus, we have the following general result.

Theorem 5 *For a pure two-qubit state written in the form $|\psi\rangle = \cos\frac{\theta}{2}|00\rangle + \sin\frac{\theta}{2}|11\rangle$ with $\theta \in [0, \frac{\pi}{2}]$, the free fraction is equal $\mu_L = \mu_F = \cos\theta$ for an arbitrary choice and number of Alice and Bob's settings.*

4 Summary

A common theme in my dissertation is entanglement, its optical generation schemes and the ensuing causal aspects. The research questions **Q1**-**Q3** provide the blueprint of research during my Doctoral studies. Let me briefly recap how those questions were answered in the respective articles.

In article [1], we construct explicit *no-touching* protocol for generating an arbitrary state of three qubits. The procedure has constant efficiency for every three-qubit state (irrespective of particles' statistics). Moreover, it requires very minimal resources: just three independent particles in the input, passive linear optics and post-selection. This result illustrates the advantage of *no-touching* designs, answering **Q1** as regards three-qubit state generation (which, in full generality, already provides enough challenge).

In article [2], we propose an explicit *no-touching* scheme for generating a multipartite W state for an arbitrary number of qubits (for both bosons and fermions). The proposal uses very minimal resources: just independent particles (without the need for auxiliary ones), passive linear optics and post-selection. We argue that it performs much better than other protocols considered in the literature (both efficiency and resource-wise). This result illustrates the advantage of the *no-touching* designs, answering **Q1** as regards multi-qubit entanglement generation (for the well-studied example of the W state).

In article [3], we investigate possible causal influences in the *quantum three box paradox*, which is a prototype of the pre- and post-selection protocols. We were interested in various types of measurement disturbances and their propagation in the experiment. In particular, it was shown using Pearl's instrumental inequalities that *parameter dependence* is necessary & enough to explain the observed statistics. We argue that the mystery of the quantum three box paradox can be dispelled since it can be seen as merely a case of the *selection bias* (or *Berkson's paradox*). The purpose of this paper was to illustrate the dangers of drawing hasty conclusions from experiments involving post-selection. This motivates **Q2** as regards the *no-touching* protocols where post-selection is an integral part of the procedure.

In article [4], we take the warning about post-selection seriously and revisit Bell's theorem with regard to the potential risk of the *selection bias*. We formulate a simple criterion, the so-called *all-but-one principle*, which is proved to safeguard the causal conclusions from the violation of Bell inequalities that are drawn using post-selected data. This explicitly answers **Q2** since the principle is fulfilled in the *no-touching* protocols. Notably, the proposed *all-but-one principle* is simple enough to be applicable in a broad range of experimental scenarios (cf., protocols with coincidence-type post-selection with the conserved number of particles).

In article [5], we assess the causal cost of violation of Bell inequalities. The comparison is based on the natural concept of *fractional measures* which ask about the frequency of violation of a given assumption (i.e., the presence of certain causal arrows) which is enough to explain the observed statistics. We show that

such defined measures of *locality* and *freedom of choice* are the same for Bell-type scenarios. In this respect, the causal framework does not prefer one assumption over another. Furthermore, we explicitly calculate those measures for the standard case of the bipartite, two-setting and two-outcome Bell scenario. Remarkably, this result gives a simple and natural interpretation of the amount of violation of the CHSH inequalities. We also analyse the case of quantum mechanical statistics. In particular, for the number of settings increasing both measures drop to zero for the maximally entangled state. This gives a broad discussion of **Q3** for general as well as quantum mechanical statistics.

The results published in Refs. [1–5] conclude the scope of this dissertation, which is organised around questions **Q1**-**Q3**. The research topics follow the natural order of increasing generality and foundational interest. Namely, I start with the construction of practical schemes of entanglement generation in passive linear optical designs (within the so-called *no-touching* paradigm). Then, I investigate the problem of *post-selection* in protocols used for the purpose of fundamental tests. Finally, I ask about the *comparison* and *weight* of causal assumptions in Bell experiments. This should give a coherent account of my scientific interest and research during my Doctoral studies.

References

[1] P. Blasiak, E. Borsuk, and M. Markiewicz. Arbitrary entanglement of three qubits via linear optics. *Sci. Rep.*, 12:21596, 2022.

[2] P. Blasiak, E. Borsuk, M. Markiewicz, and Y.-S. Kim. Efficient linear-optical generation of a multipartite W state. *Phys. Rev. A*, 104:023701, 2021.

[3] P. Blasiak and E. Borsuk. Causal reappraisal of the quantum three-box paradox. *Phys. Rev. A*, 105:012207, 2022.

[4] P. Blasiak, E. Borsuk, and M. Markiewicz. On safe post-selection for Bell tests with ideal detectors: Causal diagram approach. *Quantum*, 5:575, 2021.

[5] P. Blasiak, E. M. Pothos, J. M. Yearsley, C. Gallus, and E. Borsuk. Violations of locality and free choice are equivalent resources in Bell experiments. *Proc. Natl. Acad. Sci. USA*, 118:e2020569118, 2021. ***EurekAlert!***

[6] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki. Quantum entanglement. *Rev. Mod. Phys.*, 81:865–942, 2009.

[7] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden. Quantum cryptography. *Rev. Mod. Phys.*, 74:145, 2002.

[8] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[9] G. Tóth and I. Apellaniz. Quantum metrology from a quantum information science perspective. *J. Phys. A: Math. Theor.*, 47:424006, 2014.

[10] J. S. Bell. *Speakable and unspeakable in quantum mechanics*. Cambridge University Press, 1987.

[11] T. Norsen. *Foundations of Quantum Mechanics: An Exploration of the Physical Meaning of Quantum Theory*. Undergraduate Lecture Notes in Physics. Springer, 2017.

[12] F. Laloë. *Do we really understand quantum mechanics?* Cambridge University Press, 2012.

[13] T. Maudlin. *Philosophy of Physics: Quantum Theory*. Princeton University Press, 2019.

[14] J.-W. Pan, Z.-B. Chen, C.-Y. Lu, H. Weinfurter, A. Zeilinger, and M. Żukowski. Multiphoton entanglement and interferometry. *Rev. Mod. Phys.*, 84:777–838, 2012.

[15] M. Erhard, M. Krenn, and A. Zeilinger. Advances in high-dimensional quantum entanglement. *Nat. Rev. Phys.*, 2:365, 2020.

[16] J. Wang, F. Sciarrino, A. Laing, and M. G. Thompson. Integrated photonic quantum technologies. *Nat. Photonics*, 14:273, 2020.

[17] M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani. Experimental Realization of Any Discrete Unitary Operator. *Phys. Rev. Lett.*, 73:58, 1994.

[18] P. Migdał, J. Rodríguez-Laguna, M. Oszmaniec, and M. Lewenstein. Multiphoton states related via linear optics. *Phys. Rev. A*, 89:062329, 2014.

[19] W. Dür, G. Vidal, and J. I. Cirac. Three qubits can be entangled in two inequivalent ways. *Phys. Rev. A*, 62:062314, 2000.

[20] P. Blasiak and M. Markiewicz. Entangling three qubits without ever touching. *Sci. Rep.*, 9:20131, 2019. ***EurekAlert!***

[21] B. Yurke and D. Stoler. Bell's-inequality experiments using independent-particle sources. *Phys. Rev. A*, 46:2229, 1992.

[22] B. Yurke and D. Stoler. Einstein-Podolsky-Rosen Effects from Independent Particle Sources. *Phys. Rev. Lett.*, 68:1251, 1992.

[23] I. Neder, N. Ofek, Y. Chung, M. Heiblum, D. Mahalu, and V. Umansky. Interference between two indistinguishable electrons from independent sources. *Nature*, 448:333, 2007.

[24] S. Bose and D. Home. Duality in Entanglement Enabling a Test of Quantum Indistinguishability Unaffected by Interactions. *Phys. Rev. Lett.*, 110:140404, 2013.

[25] R. Lo Franco and G. Compagno. Quantum entanglement of identical particles by standard information-theoretic notions. *Sci. Rep.*, 6:20603, 2016.

[26] R. Lo Franco and G. Compagno. Indistinguishability of Elementary Systems as a Resource for Quantum Information Processing. *Phys. Rev. Lett.*, 120:240403, 2018.

[27] M. C. Tichy. Interference of identical particles from entanglement to boson-sampling. *J. Phys. B: At. Mol. Opt. Phys.*, 47:103001 (34pp), 2014.

[28] A. Acín, A. Andrianov, L. Costa, E. Jané, J. I. Latorre, and R. Tarrach. Generalized schmidt decomposition and classification of three-quantum-bit states. *Phys. Rev. Lett.*, 85:1560, 2000.

[29] H. A. Carteret, A. Higuchi, and A. Sudbery. Multipartite generalization of the schmidt decomposition. *Journal of Mathematical Physics*, 41(12):7932–7939, dec 2000.

[30] Y.-S. Kim, Y.-W. Cho, H.-T. Lim, and S.-W. Han. Efficient linear optical generation of a multipartite W state via a quantum eraser. *Phys. Rev. A*, 101:022337, 2020.

[31] X.-P. Zang, M. Yang, F. Ozaydin, W. Song, and Z.-L. Cao. Deterministic generation of large scale atomic W states. *Opt. Express*, 24:12293, 2016.

[32] N.-N. Wu and M. Jiang. A highly efficient scheme for joint remote preparation of multi-qubit W state with minimum quantum resource. *Quantum Inf. Process.*, 17:340, 2018.

[33] S. Bugu, F. Ozaydin, and T. Ferrus, T.and Kodera. Preparing Multipartite Entangled Spin Qubits via Pauli Spin Blockade. *Sci. Rep.*, 10:3481, 2020.

[34] B. Bellomo, R. Lo Franco, and G. Compagno. N identical particles and one particle to entangle them all. *Phys. Rev. A*, 96:022319, 2017.

[35] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner. Bell nonlocality. *Rev. Mod. Phys.*, 86:419, 2014.

[36] V. Scarani. *Bell Nonlocality*. Oxford University Press, 2019.

[37] A. Aspect, J. Dalibard, and G. Roger. Experimental Test of Bell's Inequalities Using Time-Varying Analyzers. *Phys. Rev. Lett.*, 49:1804, 1982.

[38] B. Hensen, H. Bernien, A. E. Dreau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenberg, R. F. L. Vermeulen, R. N. Schouten, C. Abellan, W. Amaya, V. Pruneri, M. W. Mitchell, M. Markham, D. J. Twitchen, D. Elkouss, S. Wehner, T. H. Taminiau, and R. Hanson. Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres. *Nature*, 526:682, 2015.

[39] M. Giustina, M. A. M. Versteegh, S. Wengerowsky, J. Handsteiner, A. Hochrainer, K. Phelan, F. Steinlechner, J. Kofler, J.-A. Larsson, C. Abellán, W. Amaya, V. Pruneri, M. W. Mitchell, J. Beyer, T. Gerrits, A. E. Lita, L. K. Shalm, S. W. Nam, T. Scheidl, R. Ursin, B. Wittmann, and A. Zeilinger. Significant-Loophole-Free Test of Bell's Theorem with Entangled Photons. *Phys. Rev. Lett.*, 115:250401, 2015.

[40] L. K. Shalm, E. Meyer-Scott, B. G. Christensen, P. Bierhorst, M. A. Wayne, M. J. Stevens, T. Gerrits, S. Glancy, D. R. Hamel, M. S. Allman, K. J. Coakley, S. D. Dyer, C. Hodge, A. E. Lita, V. B. Verma, C. Lambrocco, E. Tortorici, A. L. Migdall, Y. Zhang, D. R. Kumor, W. H. Farr, F. Marsili, M. D. Shaw, J. A. Stern, C. Abellán, W. Amaya, V. Pruneri, T. Jennewein, M. W. Mitchell, P. G. Kwiat, J. C. Bienfang, R. P. Mirin, E. Knill, and S. W. Nam. Strong Loophole-Free Test of Local Realism. *Phys. Rev. Lett.*, 115:250402, 2015.

[41] A. Aspect. Closing the Door on Einstein and Bohr's Quantum Debate. *Physics*, 8(123), 2015.

[42] D. Rauch, J. Handsteiner, A. Hochrainer, J. Gallicchio, A. S. Friedman, C. Leung, B. Liu, L. Bulla, S. Ecker, F. Steinlechner, R. Ursin, B. Hu, D. Leon, C. Benn, A. Ghedina, M. Cecconi, A. H. Guth, D. I. Kaiser, T. Scheidl, and

A. Zeilinger. Cosmic Bell Test Using Random Measurement Settings from High-Redshift Quasars. *Phys. Rev. Lett.*, 121:080403, 2018.

[43] The BIG Bell test Collaboration. Challenging local realism with human choices. *Nature*, 557:212, 2018.

[44] J.-A. Larsson. Loopholes in Bell inequality tests of local realism. *J. Phys. A: Math. Gen.*, 47:424003, 2014.

[45] J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2nd edition, 2009.

[46] P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search*. The MIT Press, 2000.

[47] J. Pearl, M. Glymour, and N. P. Jewell. *Causal Inference in Statistics: A Primer*. Wiley, 2016.

[48] M. A. Hernan and J. M. Robins. *Causal Inference: What If*. Chapman & Hall/CRC, 2020.

[49] C. H. Brans. Bell's theorem does not eliminate fully causal hidden variables. *Int. J. Theor. Phys.*, 27:219, 1988.

[50] G. 't Hooft. *The Cellular Automaton Interpretation of Quantum Mechanics*. Fundamental Theoriesof Physics, Vol. 185. Springer, 2016.

[51] S. Hossenfelder and T. Palmer. Rethinking Superdeterminism. *Front. Phys.*, 8:139, 2020.

[52] H. Price. *Time's Arrow and Archimedes' Point: New Directions for the Physics of Time*. Oxford University Press, 1996.

[53] K. B. Wharton and N. Argaman. Colloquium: Bell's theorem and locally mediated reformulations of quantum mechanics. *Rev. Mod. Phys.*, 92:021002, 2020.

[54] M. S. Leifer and M. F. Pusey. Is a time symmetric interpretation of quantum theory possible without retrocausality? *Proc. R. Soc. A*, 473:20160607, 2017.

[55] Emily Adlam. Two roads to retrocausality. *Synthese*, 200:422, 2022.

[56] Y. Aharonov and L. Vaidman. Complete description of a quantum system at a given time. *J. Phys. A: Math. Gen.*, 24:2315, 1991.

[57] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. Proposed Experiment to Test Local Hidden-Variable Theories. *Phys. Rev. Lett.*, 23:880–884, 1969.

[58] A. Fine. Hidden Variables, Joint Probability, and the Bell Inequalities. *Phys. Rev. Lett.*, 48:291–295, 1982.

[59] B. S. Tsirelson. Quantum generalizations of Bell's inequality. *Lett. Math. Phys.*, 4:93–100, 1980.

[60] J. Pearl and D. Mackenzie. *The Book of Why: The New Science of Cause and Effect*. Basic Books, 2018.

[61] M. A. Hernán, A. Alonso, and G. Logroscino. Cigarette smoking and dementia: Potential selection bias in the elderly. *Epidemiology*, 19:448, 2008.

[62] H. R. Banack and J. S. Kaufman. Does selection bias explain the obesity paradox among individuals with cardiovascular disease? *Annals of Epidemiology*, 25:342, 2015.

[63] Y. Aharonov, D. Z. Albert, and L. Vaidman. How the result of a measurement of a component of the spin of a spin-1/2 particle can turn out to be 100. *Phys. Rev. Lett.*, 60:1351–1354, 1988.

[64] R. E. Kastner. The Three-Box "Paradox"and Other Reasons to Reject the Counterfactual Usage of the ABL Rule. *Found. Phys.*, 29:851, 1999.

[65] L. Vaidman. The Meaning of Elements of Reality and Quantum Counterfactuals: Reply to Kastner. *Found. Phys.*, 29:865, 1999.

[66] J. Finkelstein. What is paradoxical about the "Three-box paradox"? *arXiv:quant-ph/0606218*, 2006.

[67] K. A. Kirkpatrick. Classical three-box 'paradox'. *J. Phys. A: Math. Gen.*, 36:4891, 2003.

[68] M. S. Leifer and R. W. Spekkens. Pre- and Post-Selection Paradoxes and Contextuality in Quantum Mechanics. *Phys. Rev. Lett.*, 95:200405, 2005.

[69] T. Ravon and L. Vaidman. The three-box paradox revisited. *J. Phys. A: Math. Gen.*, 40:2873, 2007.

[70] K. A. Kirkpatrick. Reply to 'The three-box paradox revisited' by T Ravon and L Vaidman. *J. Phys. A: Math. Gen.*, 40:2883, 2007.

[71] O. J. E. Maroney. Measurements, disturbances and the quantum three box paradox. *Stud. Hist. Phil. Mod. Phys.*, 58:41, 2017.

[72] Y. Aharonov, P. G. Bergmann, and J. L. Lebowitz. Time Symmetry in the Quantum Process of Measurement. *Phys. Rev.*, 134:B1410, 1964.

[73] A. Peres. *Quantum Theory: Concepts and Methods*. Kluwer Academic Publishers, 1995.

[74] J. Kysela, M. Erhard, A. Hochrainer, M. Krenn, and A. Zeilinger. Path identity as a source of high-dimensional entanglement. *Proc. Natl. Acad. Sci. U.S.A.*, 117:26118, 2020.

[75] M. Krenn, A. Hochrainer, M. Lahiri, and A. Zeilinger. Entanglement by Path Identity. *Phys. Rev. Lett.*, 118:080401, 2017.

[76] K. Tschernig, C. Müller, M. Smoor, T. Kroh, J. Wolters, O. Benson, K. Busch, and A. Perez-Leija. Direct observation of the particle exchange phase of photons. *Nat. Photon.*, 15:671, 2021.

[77] H.-S. Zhong, Y. Li, W. Li, L.-C. Peng, Z.-E. Su, Y. Hu, Y.-M. He, X. Ding, W. Zhang, H. Li, L. Zhang, Z. Wang, L. You, X.-L. Wang, X. Jiang, L. Li, Y.-A. Chen, N.-L. Liu, C.-Y. Lu, and J.-W. Pan. 12-Photon entanglement and Scalable Scattershot Boson Sampling with Optimal Entangled-Photon Pairs from Parametric Down-Conversion. *Phys. Rev. Lett.*, 121:250505, 2018.

[78] A. C. Elitzur, S. Popescu, and D. Rohrlich. Quantum nonlocality for each pair in an ensemble. *Phys. Lett. A*, 162:25, 1992.

[79] L. Hardy. A new way to obtain Bell inequalities. *Phys. Lett. A*, 161:21, 1991.

[80] J. Barrett, A. Kent, and S. Pironio. Maximally Nonlocal and Monogamous Quantum Correlations. *Phys. Rev. Lett.*, 97:170409, 2006.

[81] R. Colbeck and R. Renner. Hidden Variable Models for Quantum Theory Cannot Have Any Local Part. *Phys. Rev. Lett.*, 101:050403, 2008.

[82] R. Colbeck and R. Renner. The Completeness of Quantum Theory for Predicting Measurement Outcomes. In G. Chiribella and R. W. Spekkens, editors, *Quantum Theory: Informational Foundations and Foils*, pages 497–528. Springer, 2016.

[83] S. Portmann, C. Branciard, and N. Gisin. Local content of all pure two-qubit states. *Phys. Rev. A*, 86:012104, 2012.

[84] M. J. W. Hall. Local Deterministic Model of Singlet State Correlations Based on Relaxing Measurement Independence. *Phys. Rev. Lett.*, 105:250404, 2010.

[85] M. J. W. Hall. Relaxed Bell inequalities and Kochen-Specker theorems. *Phys. Rev. A*, 84:022102, 2011.

[86] J. Barrett and N. Gisin. How Much Measurement Independence Is Needed to Demonstrate Nonlocality? *Phys. Rev. Lett.*, 106:100406, 2011.

[87] M. J. W. Hall and C. Branciard. Measurement-dependence cost for Bell nonlocality: Causal versus retrocausal models. *Phys. Rev. A*, 102:052228, 2020.

[88] G. Pütz, D. Rosset, T. J. Barnea, Y.-C. Liang, and N. Gisin. Arbitrarily Small Amount of Measurement Independence Is Sufficient to Manifest Quantum Nonlocality. *Phys. Rev. Lett.*, 113:190402, 2014.

[89] D. Aktas, S. Tanzilli, A. Martin, G. Pütz, R. Thew, and N. Gisin. Demonstration of Quantum Nonlocality in the Presence of Measurement Dependence. *Phys. Rev. Lett.*, 114:220404, 2015.

[90] G. Pütz and N. Gisin. Measurement dependent locality. *New J. Phys.*, 18:055006, 2016.

5 Articles

The following articles [1–5] contain all results presented in my PhD thesis:

[1] P. Blasiak, E. Borsuk and M. Markiewicz.
Arbitrary entanglement of three qubits via linear optics.
Scientific Reports (Nature) **12** 21596 (2022).

[2] P. Blasiak, E. Borsuk, M. Markiewicz and Y.-S. Kim.
Efficient linear optical generation of a multipartite W state.
Physical Review A **104** 023701 (2021).

[3] P. Blasiak and E. Borsuk.
Causal reappraisal of the quantum three-box paradox.
Physical Review A **105** 012207 (2022).

[4] P. Blasiak, E. Borsuk, and M. Markiewicz.
On safe post-selection for Bell nonlocality: Causal diagram approach.
Quantum **5** 575 (2021).

[5] P. Blasiak, E. M. Pothos, J. M. Yearsley, C. Gallus, and E. Borsuk.
Violations of locality and free choice are equivalent resources in Bell experiments.
Proceedings of the National Academy of Sciences (PNAS) **118** e2020569118 (2021).

EurekAlert! (Press note May 2021):
We know the cost of free choice and locality - in physics and not only.

scientific reports

Check for updates

OPEN

Arbitrary entanglement of three qubits via linear optics

Pawel Blasiak[1,2]✉, Ewa Borsuk[2] & Marcin Markiewicz[3]

We present a linear-optical scheme for generating an arbitrary state of three qubits. It requires only three independent particles in the input and post-selection of the coincidence type at the output. The success probability of the protocol is equal for any desired state. Furthermore, the optical design remains insensitive to particle statistics (bosons, fermions or anyons). This approach builds upon the no-touching paradigm, which demonstrates the utility of particle indistinguishability as a resource of entanglement for practical applications.

Entanglement remains a central theme in quantum foundations research[1,2]. It is considered a key resource enabling the advantage in quantum information tasks[3,4]. There is therefore a vital interest in practical entanglement generation schemes capable of delivering the broadest possible range of states, from which one might choose a state desired for a problem at hand. Ideally, we would like to have a generic and efficient tool for constructing an arbitrary multi-particle state from some simpler initial state (possibly having just a few independent particles to begin with). A paradigmatic example is the construction of the full class of single-particle states (a qudit). In this case, arbitrary unitary transformation can be experimentally implemented by linear-optical devices[5] and hence any single-particle state can be deterministically prepared from any given initial state. However, it is not true for multi-particle states that do not transform one into another by means of linear optics[6]. Thus for the generation of multi-partite entanglement, non-linear effects or post-selection of some sort must be employed. Several techniques have been developed to this effect which produce certain classes of states[7–10]. However, there is no systematic linear-optical method for obtaining an arbitrary multi-particle entangled state that would start with a supply of independent particles in the input.

In this work, we focus on entanglement generation for three qubits. Notably, this is the first non-trivial case where the arbitrariness of the desired state becomes challenging. An interesting approach to this problem consists in considering the set of stochastic local operations and classical communication (SLOCC)[11–13]. It has been shown that for three qubits there are two inequivalent classes of genuinely tripartite entangled states[12]. This means that using SLOCCs for filtering arbitrary tripartite entanglement requires two different types of entangled states to begin with, i.e., states of the GHZ and the W type. However, it should be remarked that the efficiency of such protocols drops to zero when moving away from the initial state. Another important result concerns the full set of linear optical transformations (without post-selection)[6]. Then the situation is further complicated as for three qubits the set of entangled states splits into a continuous number of inequivalent classes. Those results illustrate the difficulties in efficiently generating arbitrary entanglement using linear operations without any entanglement from the outset.

Here we will show that linear-optical transformations augmented with post-selection of the coincidence type are enough to generate arbitrary entanglement of three qubits from three independent particles (i.e., without requiring any prior entanglement). Our proposal builds on the so-called no-touching paradigm in optical designs which draws from the inherent indistinguishability of particles; see Ref.[14] for a general scheme and Refs.[15–19] for some examples. Notably, the protocol is an instance of direct and explicit construction of any given state. A distinctive advantage of the proposal is that it has equal efficiency for generating any desired state and it is insensitive to particle statistics.

Results

Simplification by generalised Schmidt decomposition. A general state of three qubits reads

$$|\psi\rangle = \sum_{ijk=0,1} \alpha_{ijk} |ijk\rangle \,, \tag{1}$$

[1]Institute for Quantum Studies, Chapman University, Orange, CA 92866, USA. [2]Institute of Nuclear Physics Polish Academy of Sciences, 31342 Kraków, Poland. [3]International Centre for Theory of Quantum Technologies, University of Gdańsk, 80308 Gdańsk, Poland. ✉email: pawel.blasiak@ifj.edu.pl

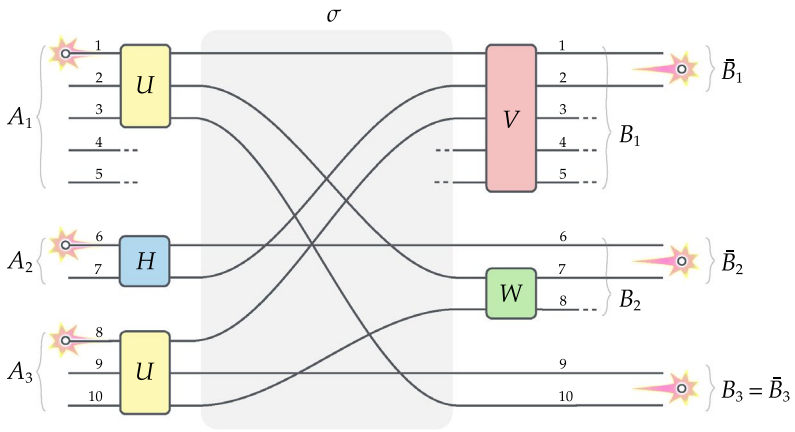


Figure 1. No-touching design for arbitrary state of three qubits. Three independent identical particles entering the optical circuit undergo a sequence of transformations which consists of local unitaries U, H and U on subsystems A_1, A_2 and A_3, followed by permutation of the paths σ, and again local unitaries V and W on subsystems B_1 and B_2. Post-selection on a single particle in each of the dual-rail qubits generates an arbitrary state of three qubits $\bar{B}_1$, $\bar{B}_2$ and $\bar{B}_3$ in Eq. (2) by the appropriate choice of unitaries in Eq. (6) as specified in Eqs. (18)–(20).

where $|ijk\rangle$ is a computational basis in $\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2$. It turns out that this can be simplified by local transformations to the combination of the following five states

$$|\psi\rangle = a\,|000\rangle + b\,e^{i\varphi}\,|100\rangle + c\,|110\rangle + d\,|101\rangle + e\,|111\rangle\,, \tag{2}$$

with five real parameters $a, b, c, d, e \geq 0$ and one phase $0 \leq \varphi \leq \pi$, such that

$$a^2 + b^2 + c^2 + d^2 + e^2 = 1. \tag{3}$$

This follows from the generalised Schmidt decomposition for three qubits[20,21] and no further reduction of the number of non-vanishing terms is possible.

Therefore, to generate the state in Eq. (1), it is enough to provide the state in Eq. (2) and then make a local transformation on each qubit. In the following, we give the explicit protocol for optical construction of an arbitrary state in the reduced form of Eq. (2).

Optical realisation. Consider three identical particles injected into an optical circuit which consists of 10 paths (or modes). The particle statistics (bosons, fermions or anyons) is irrelevant for the purpose at hand. Following the idea presented in Ref.[14] we will group paths at the input and output into three systems, see Fig. 1:

$$\begin{aligned} &\textit{Input:}\ A_1 = \{1,2,3,4,5\},\ A_2 = \{6,7\},\ A_3 = \{8,9,10\},\\ &\textit{Output:}\ B_1 = \{1,2,3,4,5\},\ B_2 = \{6,7,8\},\ B_3 = \{9,10\}. \end{aligned}$$

Furthermore, we will distinguish three pairs of paths $\bar{B}_1$, $\bar{B}_2$ and $\bar{B}_3$ in the respective subsystems at the output, i.e. $\bar{B}_k \subset B_k$. They will play the role of the so-called *dual-rail qubits*, where the computational basis $|0\rangle, |1\rangle$ is encoded by the presence of a single particle in the respective path of a given pair $\bar{B}_k$. Accordingly, we have the following representation for a general qubit state $\alpha\,|0\rangle + \beta\,|1\rangle$:

$$\begin{aligned} \bar{B}_1 = \{1,2\} \quad &\rightsquigarrow \textit{Qubit 1:}\ (\alpha\,a_1^\dagger + \beta\,a_2^\dagger)\,|\Omega\rangle\,,\\ \bar{B}_2 = \{6,7\} \quad &\rightsquigarrow \textit{Qubit 2:}\ (\alpha\,a_6^\dagger + \beta\,a_7^\dagger)\,|\Omega\rangle\,,\\ \bar{B}_3 = \{9,10\} \quad &\rightsquigarrow \textit{Qubit 3:}\ (\alpha\,a_9^\dagger + \beta\,a_{10}^\dagger)\,|\Omega\rangle\,. \end{aligned}$$

where $|\Omega\rangle$ denotes the vacuum state and $a_i^\dagger$ are the usual particle creation operators in the second quantisation formalism. We note that the dual-rail encoding assumes the presence of a *single* particle in a given pair of paths $\bar{B}_k$ for $k = 1, 2, 3$. In our scheme this will be guaranteed by *post-selection* on a single particle in each dual-rail qubit $\bar{B}_k$.

The optical protocol is depicted in Fig. 1. It consists of a sequence of three unitary transformations on three independent particles injected in paths 1, 6 and 8. First, the particles undergo local unitaries U, H and U in each subsystem A_1, A_2 and A_3. Second, the paths are rearranged according to some permutation $\sigma \in \mathcal{S}_{10}$. Third, there are two local unitaries V and W on subsystems B_1 and B_2 implemented at the output. Finally, the protocol ends with post-selection on a single particle in each pair of paths $\bar{B}_1$, $\bar{B}_2$ and $\bar{B}_3$ which generates three dual-rail encoded qubits.

Now, we can make the unitaries in Fig. 1 more precise. Let the first two transformations U and V produce symmetric superposition of the injected particles, which in the matrix notation amounts to

$$U = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 & . & . \\ 1 & . & . \\ 1 & . & . \end{pmatrix} \quad \text{and} \quad H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & . \\ 1 & . \end{pmatrix}. \tag{4}$$

The permutation of modes $\sigma \in \mathcal{S}_{10}$ is given as follows

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 1 & 7 & 10 & 4 & 5 & 6 & 2 & 3 & 9 & 8 \end{pmatrix}. \tag{5}$$

The final unitaries V and W are defined in a non-trivial way by the following two matrices

$$V = \begin{pmatrix} \kappa & 0 & 0 & \delta & \epsilon \\ \bar{\delta} & \mu & \nu & -\bar{\kappa} & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \end{pmatrix} \quad \text{and} \quad W = \begin{pmatrix} \xi & \tau \\ . & . \end{pmatrix}, \tag{6}$$

with some parameters $\kappa, \delta, \nu, \mu, \epsilon, \xi$ and τ. In the above notation, the dots "." are left unspecified and chosen so that the matrices are unitary. Observe that this can always be done by augmenting the missing columns/rows to an orthonormal basis (note that the two upper rows of V are orthogonal at the outset). The only constraint on the parameters $\kappa, \delta, \nu, \mu, \epsilon, \xi$ and τ is their respective normalisation, i.e.

$$|\xi|^2 + |\tau|^2 = 1, \tag{7}$$

$$|\kappa|^2 + |\delta|^2 + |\epsilon|^2 = 1, \tag{8}$$

$$|\delta|^2 + |\mu|^2 + |\nu|^2 + |\kappa|^2 = 1. \tag{9}$$

For our purposes, the dotted entries "." will play no role in the argument (in the following, they contribute only to the terms that drop out upon post-selection). All the remaining parameters $\kappa, \delta, \nu, \mu, \epsilon, \xi$ and τ will be specified shortly.

Let us write out the state that results from the protocol in Fig. 1 after injecting three independent particles in paths 1, 6, 8 and post-selecting on a single particle in each dual-rail qubit $\bar{B}_1$, $\bar{B}_2$ and $\bar{B}_3$. The evolution of the system is given by a sequence of steps as described in the following lines:

$$a_1^\dagger a_6^\dagger a_8^\dagger |\Omega\rangle \xrightarrow[Eq.(4)]{U,H,U} \frac{1}{3\sqrt{2}}\left(a_1^\dagger + a_2^\dagger + a_3^\dagger\right)\left(a_6^\dagger + a_7^\dagger\right)\left(a_8^\dagger + a_9^\dagger + a_{10}^\dagger\right)|\Omega\rangle \tag{10}$$

$$\xrightarrow[Eq.(5)]{\sigma} \frac{1}{3\sqrt{2}}\left(a_1^\dagger + a_7^\dagger + a_{10}^\dagger\right)\left(a_6^\dagger + a_2^\dagger\right)\left(a_3^\dagger + a_9^\dagger + a_8^\dagger\right)|\Omega\rangle \tag{11}$$

$$\xrightarrow[Eq.(6)]{V,W} \frac{1}{3\sqrt{2}}\Big(\left(\kappa a_1^\dagger + \bar{\delta} a_2^\dagger + ...\right) + \left(\xi a_7^\dagger + ...\right) + a_{10}^\dagger\Big) \tag{12}$$

$$\Big(a_6^\dagger + \left(\mu\, a_2^\dagger + \cdots\right)\Big)\Big(\left(\nu\, a_2^\dagger + \cdots\right) + a_9^\dagger + \left(\tau\, a_7^\dagger + \cdots\right)\Big)|\Omega\rangle \tag{13}$$

$$\overset{post\text{-}select}{\rightsquigarrow} \frac{1}{3\sqrt{2}}(\kappa a_1^\dagger a_6^\dagger a_9^\dagger + \bar{\delta} a_2^\dagger a_6^\dagger a_9^\dagger + \xi\mu a_7^\dagger a_2^\dagger a_9^\dagger + \nu a_{10}^\dagger a_6^\dagger a_2^\dagger + \mu\tau a_{10}^\dagger a_2^\dagger a_7^\dagger)|\Omega\rangle \tag{14}$$

$$= \frac{1}{3\sqrt{2}}\Big(\kappa\,|000\rangle + \bar{\delta}\,|100\rangle + \xi\mu\,|110\rangle + \nu\,|101\rangle + \mu\tau\,|111\rangle\Big), \tag{15}$$

where the last equality holds for bosons in the dual-rail encoding. This renders the desired state in Eq. (2) when

$$\kappa = a, \;\; \bar{\delta} = b\, e^{i\varphi}, \;\; \xi\mu = c, \;\; \nu = d, \;\; \mu\tau = e. \tag{16}$$

We observe that we can always choose matrices W and V to satisfy these equations by defining parameters $\kappa, \delta, \nu, \mu, \epsilon, \xi$ and τ in the following way

$$\kappa = a, \qquad \delta = b\, e^{-i\varphi}, \qquad \nu = d, \tag{17}$$

$$\mu = \sqrt{c^2 + e^2}, \qquad \epsilon = \sqrt{1 - a^2 - b^2}, \tag{18}$$

$$\xi = \frac{c}{\mu}, \qquad \tau = \frac{e}{\mu}. \tag{19}$$

(If $\mu = 0$, then ξ and τ can be taken arbitrarily). It is straightforward to check that the conditions in Eqs. (7)–(9) are satisfied, since the constraint in Eq. (3) holds.

Note that the scheme works for fermions as well. In this case, the final state in Eq. (16) takes the form

$$\kappa\,|000\rangle + \bar{\delta}\,|100\rangle - \xi\mu\,|110\rangle - \nu\,|101\rangle + \mu\tau\,|111\rangle\,, \tag{20}$$

and then the Eqs. (18)–(20) require a trivial modification $\nu = -d$ and $\xi = -\frac{c}{\mu}$ in order to recover the desired state in Eq. (2). In a similar manner it is straightforward to adjust Eq. (16) for *any* particle statistics (anyons).

Finally, we observe that the expression in Eq. (16) is unnormalised due to post-selection. This allows to read off the success probability (efficiency) of the process which is equal to $\left(\frac{1}{3\sqrt{2}}\right)^2 = \frac{1}{18} \approx 5\%$. Notably, the efficiency is the *same* for *every* three-qubit state $|\psi\rangle$.

Discussion

We remark that the above-described protocol, based on dual-rail encoded qubits, provides a ready-made template that straightforwardly translates into any other physical implementation of qubits. This turns out to be a generic feature of the no-touching designs in which the question of particle statistics becomes virtually irrelevant because of post-selection. It is due to the fact that the latter projects on the sector where the evolution of the system features at most a single particle in each mode, which makes immaterial the distinction between the bunching and anti-bunching effects for bosons and fermions; see Refs.[14–19] for a discussion.

From the fundamental point of view, it is interesting to note the significance of the inherent indistinguishability of particles as conveniently described in the second quantisation formalism. It appears that entanglement resulting from the symmetrization postulate can be treated as a genuine resource and transformed into other kinds of entanglement which can be directly observed and used for practical applications[14,22–24]. This paper shows that arbitrary entanglement of three qubits can be extracted in this way. For an extension to some multi-particle entangled states see e.g. Refs.[19,25–29].

An important advantage of the proposed protocol is the minimal amount of resources employed to generate an arbitrary three-qubit state compared to the existing techniques:

(a) It requires *only* linear optics and works equally well for *any* particle statistics, cf. Refs.[7–10].
(b) There is *no* need for auxiliary systems (particles) or measurements, cf. Refs.[25,27].
(c) The protocol requires *only* three independent particles in the input, i.e., no prior entanglement is required.
(d) It has the *same* efficiency for the generation of *any* desired three-qubit state.

This distinguishes our proposal from the typical approach based on filtering via SLOCC operations which requires auxiliary entanglement from the outset, and its success probability for arbitrary three-qubit states drops to zero; see **Methods** section for discussion. Moreover, the generation of states via SLOCC filtering generally demands different initial states depending on the SLOCC equivalence class of the target state. For optical proposals aimed at preparation of single representatives in the SLOCC classes for the purpose of filtering see Refs.[14,26,28,30]. Notably, our protocol overcomes the division into SLOCC equivalence classes due to the presence of mode permutation σ, which is a non-local operation from the point of view of subsystems defined by mode grouping.

We note that our scheme relies on a specific type of post-selection which requires *coincidence* count in the output channels encoding dual-rail qubits $\bar{B}_1$, $\bar{B}_2$ and $\bar{B}_3$. A direct way to check the coincidence criterion involves the measurement of each qubit. Typically this destroys the state, but the recorded correlations can be still used for the extraction of some information relevant to a given experiment. The utility of such a direct post-selection depends on the task at hand. For illustration see, e.g., the recent boson sampling experiments[31] or the direct verification of the boson nature of photons[32,33]. It was also shown to be safe for Bell-type tests of non-locality[24]. We remark that many modern state generation schemes refer to post-selection of the the coincidence-type, like e.g. entanglement by path identity[9,34,35] or spatial overlap of indistinguishable particles[22,23,36]. However, if the generated state needs to be further processed, then direct detection does not meet this requirement. In such a case, the solution is provided by non-demolition measurements carried out on each dual-rail qubit $\bar{B}_1$, $\bar{B}_2$ and $\bar{B}_3$. Such a measurement ascertains the presence of a particle without destroying it and not affecting its state. Therefore, a positive joint result of those three (non-demolition) measurements heralds the generation of the desired state in the signal modes. This turns the protocol into an *event-ready* scheme. Notably, a few non-demolition measurement techniques have been developed in recent years. See e.g. Refs.[37–39] for non-demolition detection of photons (noting that polarization and dual-rail encoding of qubits transforms one into another via polarizing beam splitters).

In summary, the characteristic features of the proposed scheme for state generation are marked by simplicity (just linear optics and post-selection of the coincidence type), limited initial resources (just three independent particles in the input), and universal efficiency (equal for any desired state). This makes the proposal an interesting technique for integrated quantum technologies motivating further research towards an extension to an arbitrary number of qubits; see the recent progress in the optical generation of certain multi-particle states within the no-touching paradigm[19,25–28,30]. We also indicate possible further improvements using a graph-theoretical approach to the analysis of linear optical schemes[40].

Methods

Comparison with the generation of arbitrary states via SLOCC operations. In this paper, we presented a universal interferometric protocol for generating an arbitrary three-qubit state from an input product state of three particles, which prepares an arbitrary state with constant finite efficiency. Here, we compare this scheme with the generation of arbitrary states from a GHZ class starting from the GHZ input state via SLOCC

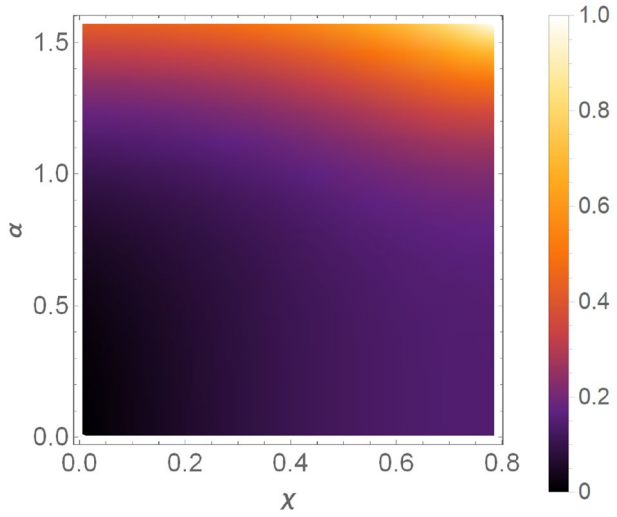


Figure 2. Success probability of obtaining arbitrary state from a two-parameter subclass of the GHZ class in Eq. (22) of the form $|\psi_{\mathrm{GHZ}}(\chi,\pi,\alpha,\alpha,\alpha)\rangle$. For $\chi=\frac{\pi}{4}$ and $\alpha\to 0$, the success probability vanishes, indicating that these states cannot be effectively obtained via SLOCC filtering.

operations. We will see that the lattr method of state generation has a vanishing efficiency for some states in this class.

As shown in the seminal paper by Dür et al.[12] an arbitrary state from the GHZ class can be parametrised by five real parameters as

$$|\psi_{\mathrm{GHZ}}(\chi,\theta,\alpha_1,\alpha_2,\alpha_3)\rangle = \sqrt{K}\big(\cos(\chi)\,|000\rangle + \sin(\chi)e^{i\theta}\,|s_1\rangle\,|s_2\rangle\,|s_3\rangle\big), \tag{21}$$

in which the normalisation constant reads $K=(1+2\cos(\chi)\sin(\chi)\cos(\alpha_1)\cos(\alpha_2)\cos(\alpha_3)\cos(\theta))^{-1}$ and the states $|s_i\rangle$ are given by $\cos(\alpha_i)\,|0\rangle+\sin(\alpha_i)\,|1\rangle$. The ranges of the parameters are as follows: $\chi\in(0,\frac{\pi}{4}]$, $\alpha_i\in(0,\frac{\pi}{2}]$ and $\theta\in[0,2\pi)$. This state can be obtained from the standard GHZ state $|\psi_{\mathrm{GHZ}}\rangle=\frac{1}{\sqrt{2}}(|000\rangle+|111\rangle)$ via SLOCC filtering operations specified by

$$|\psi_{\mathrm{GHZ}}(\chi,\theta,\alpha_1,\alpha_2,\alpha_3)\rangle = M(\chi,\theta,\alpha_1,\alpha_2,\alpha_3)\,|\psi_{\mathrm{GHZ}}\rangle\,, \tag{22}$$

where the SLOCC operator M has the form[12]

$$\begin{aligned} M(\chi,\theta,\alpha_1,\alpha_2,\alpha_3) &= \sqrt{2K}\begin{pmatrix}\cos(\chi) & \sin(\chi)\cos(\alpha_1)e^{i\theta}\\ 0 & \sin(\chi)\sin(\alpha_1)e^{i\theta}\end{pmatrix}\otimes\begin{pmatrix}1 & \cos(\alpha_2)\\ 0 & \sin(\alpha_2)\end{pmatrix}\otimes\begin{pmatrix}1 & \cos(\alpha_3)\\ 0 & \sin(\alpha_3)\end{pmatrix}\\ &= \sqrt{2K}\,\tilde{M}(\chi,\theta,\alpha_1,\alpha_2,\alpha_3).\end{aligned} \tag{23}$$

Such a filtering operation can be implemented as a two-outcome POVM measurement[13], with measurement operators defined by $P=M/||M||$ and $P'=\sqrt{\mathbb{1}-P^\dagger P}$. The outcome related to the measurement operator P indicates the success of the protocol, whereas the outcome related to P'—its failure. The norm has to be chosen to guarantee that $P^\dagger P\le\mathbb{1}$. One of the typical choices is the spectral norm of the operator M, defined as the largest singular value of M. This choice turns out to be optimal for the task of entanglement distillation of two-qubit states[13]. However, other choices that guarantee the condition $P^\dagger P\le\mathbb{1}$, such as the Frobenius norm, are also correct. The success probability of filtering arbitrary state of the form Eq. (22) for SLOCC operator M is thus given by[41]

$$p_{\mathrm{succ}} = \mathrm{Tr}(P\rho_{\mathrm{GHZ}}P^\dagger) = \mathrm{Tr}\left(\frac{M}{||M||}\rho_{\mathrm{GHZ}}\frac{M^\dagger}{||M^\dagger||}\right) = \frac{\mathrm{Tr}(M\rho_{\mathrm{GHZ}}M^\dagger)}{||M||^2} = \frac{1}{||M||^2}, \tag{24}$$

where $\rho_{\mathrm{GHZ}}=|\psi_{\mathrm{GHZ}}\rangle\langle\psi_{\mathrm{GHZ}}|$, and the last equality follows from the fact that $|\psi_{\mathrm{GHZ}}\rangle$ is already properly normalised. Note that the operator M is *not* unitary, and therefore it does not preserve the normalisation of a general state it acts on—the state ρ_{GHZ} is an exception.

Let us assume that we choose the spectral norm in Eq. (25). For the clarity of presentation we will focus on a two-parameter subclass of states from the GHZ class Eq. (22) of the form $|\psi_{\mathrm{GHZ}}(\chi,\pi,\alpha,\alpha,\alpha)\rangle$. In Fig. 2 we present the success probability of obtaining this state from the GHZ state as a function of parameters χ and α. We can see that the probability tends to zero for $\chi=\frac{\pi}{4}$ and $\alpha\to 0$, which stands in sharp contrast with our protocol that allows for the generation of these states with the fixed finite probability of success independently of the values of the parameters. One may argue that the effect of vanishing probability is related to a specific choice of the norm. However, it is easy to see that this effect holds for any choice of the norm consistent with

the condition $P^\dagger P \leq \mathbb{1}$. Indeed, it suffices to show that $||M(\frac{\pi}{4},\pi,\alpha \to 0)|| \to \infty$ for any choice of the norm. Due to Eq. (24) we have

$$||M(\tfrac{\pi}{4},\pi,\alpha \to 0)|| \;=\; \left|\sqrt{2K\left(\tfrac{\pi}{4},\pi,\alpha \to 0\right)}\right| \cdot ||\tilde{M}(\tfrac{\pi}{4},\pi,\alpha \to 0)||. \tag{25}$$

Now it can be easily verified that $\left|\sqrt{2K\left(\frac{\pi}{4},\pi,\alpha \to 0\right)}\right| \to \infty$. Therefore it suffices to show that $||\tilde{M}(\frac{\pi}{4},\pi,\alpha \to 0)||$ is strictly positive for any choice of the norm. For the spectral norm one has $||\tilde{M}(\frac{\pi}{4},\pi,\alpha \to 0)|| = 2$. However, all matrix norms for finite dimensional matrices of a fixed dimension are equivalent, which means that for any two norms $||\cdot||_X$ and $||\cdot||_Y$ there exist two *positive* numbers x, x' such that for any matrix A one has $x||A||_X \leq ||A||_Y \leq x'||A||_X$. From this property it follows that $||\tilde{M}(\frac{\pi}{4},\pi,\alpha \to 0)||$ must be strictly positive for any choice of the matrix norm, which implies $||M(\frac{\pi}{4},\pi,\alpha \to 0)|| \to \infty$ and therefore the success probability for filtering the states in the neighbourhood of $\chi = \frac{\pi}{4}$ and $\alpha = 0$, for any implementation of the SLOCC operation in Eq. (24), is arbitrarily close to zero.

This shows that our protocol overcomes the difficulties of state generation via SLOCC filtering operations, since in the latter: (i) the filtering probability can vanish, (ii) we are confined within one of the six entanglement classes depending on the initial state of the filtering. Both restrictions do not apply to our protocol, in which the success probability is constant for any state and we can reach an arbitrary state from the same trivial initial state.

Data availability

All data generated or analysed during this study are included in this published article.

Received: 22 August 2022; Accepted: 19 October 2022
Published online: 14 December 2022

References

1. Bell J. S. *Speakable and Unspeakable in Quantum Mechanics* (Cambridge University Press, 1987).
2. Brunner, N., Cavalcanti, D., Pironio, S., Scarani, V. & Wehner, S. Bell nonlocality. *Rev. Mod. Phys.* **86**, 419 (2014).
3. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge University Press, 2000).
4. Horodecki, R., Horodecki, P., Horodecki, M. & Horodecki, K. Quantum entanglement. *Rev. Mod. Phys.* **81**, 865–942 (2009).
5. Reck, M., Zeilinger, A., Bernstein, H. J. & Bertani, P. Experimental realization of any discrete unitary operator. *Phys. Rev. Lett.* **73**, 58 (1994).
6. Migdał, P., Rodríguez-Laguna, J., Oszmaniec, M. & Lewenstein, M. Multiphoton states related via linear optics. *Phys. Rev. A* **89**, 062329 (2014).
7. Pan, J.-W. *et al.* Multiphoton entanglement and interferometry. *Rev. Mod. Phys.* **84**, 777–838 (2012).
8. Krenn, M., Malik, M., Fickler, R., Lapkiewicz, R. & Zeilinger, A. Automated search for new quantum experiments. *Phys. Rev. Lett.* **116**, 090405 (2016).
9. Erhard, M., Krenn, M. & Zeilinger, A. Advances in high-dimensional quantum entanglement. *Nat. Rev. Phys.* **2**, 365 (2020).
10. Wang, J., Sciarrino, F., Laing, A. & Thompson, M. G. Integrated photonic quantum technologies. *Nat. Photonics* **14**, 273 (2020).
11. Kent, A., Linden, N. & Massar, S. Optimal entanglement enhancement for mixed states. *Phys. Rev. Lett.* **83**, 2656 (1999).
12. Dür, W., Vidal, G. & Cirac, J. I. Three qubits can be entangled in two inequivalent ways. *Phys. Rev. A* **62**, 062314 (2000).
13. Verstraete, F., Dehaene, J. & DeMoor, B. Local filtering operations on two qubits. *Phys. Rev. A* **64**, 010101 (2001).
14. Blasiak, P. & Markiewicz, M. Entangling three qubits without ever touching. *Sci. Rep.* **9**, 20131 (2019).
15. Yurke, B. & Stoler, D. Einstein–Podolsky–Rosen effects from independent particle sources. *Phys. Rev. Lett.* **68**, 1251 (1992).
16. Yurke, B. & Stoler, D. Bell's-inequality experiments using independent-particle sources. *Phys. Rev. A* **46**, 2229 (1992).
17. Neder, I. *et al.* Interference between two indistinguishable electrons from independent sources. *Nature* **448**, 333 (2007).
18. Bose, S. & Home, D. Duality in entanglement enabling a test of quantum indistinguishability unaffected by interactions. *Phys. Rev. Lett.* **110**, 140404 (2013).
19. Blasiak, P., Borsuk, E., Markiewicz, M. & Kim, Y.-S. Efficient linear-optical generation of a multipartite W state. *Phys. Rev. A* **104**, 023701 (2021).
20. Acín, A. *et al.* Generalized schmidt decomposition and classification of three-quantum-bit states. *Phys. Rev. Lett.* **85**, 1560 (2000).
21. Carteret, H. A., Higuchi, A. & Sudbery, A. Multipartite generalization of the Schmidt decomposition. *J. Math. Phys.* **41**, 7932–7939 (2000).
22. Lo Franco, R. & Compagno, G. "Quantum entanglement of identical particles by standard information-theoretic notions". *Sci. Rep.* **6**, 20603 (2016).
23. Lo Franco, R. & Compagno, G. Indistinguishability of elementary systems as a resource for quantum information processing. *Phys. Rev. Lett.* **120**, 240403 (2018).
24. Blasiak, P., Borsuk, E. & Markiewicz, M. On safe post-selection for Bell tests with ideal detectors: Causal diagram approach. *Quantum* **5**, 575 (2021).
25. Bellomo, B., Lo Franco, R. & Compagno, G. N identical particles and one particle to entangle them all. *Phys. Rev. A* **96**, 022319 (2017).
26. Kim, Y.-S. *et al.* Informationally symmetrical Bell state preparation and measurement. *Opt. Express* **26**, 29539 (2018).
27. Kim, Y.-S., Cho, Y.-W., Lim, H.-T. & Han, S.-W. Efficient linear optical generation of a multipartite W state via a quantum eraser. *Phys. Rev. A* **101**, 022337 (2020).
28. Ju, L., Yang, M., Paunković, N., Chu, W.-J. & Cao, Z.-L. Creating photonic GHZ and W states via quantum walk. *Quantum Inf. Process.* **18**, 176 (2019).
29. Borsuk, E. & Blasiak, P. Generation of arbitrary Dicke states in a linear optical circuit (2022) (**in preparation**).
30. Lee, D. *et al.* Entangling three identical particles via spatial overlap. *Opt. Express* **30**, 30525 (2022).
31. Wang, H. *et al.* Boson sampling with 20 input photons and a 60-mode interferometer in a 10^{14}-dimensional Hilbert space demonstration. *Phys. Rev. Lett.* **123**, 250503 (2019).
32. Tschernig, K. *et al.* Direct observation of the particle exchange phase of photons. *Nat. Photon.* **15**, 671 (2021).
33. Lo Franco, R. Directly proving the bosonic nature of photons. *Nat. Photon.* **15**, 638 (2021).
34. Krenn, M., Hochrainer, A., Lahiri, M. & Zeilinger, A. Entanglement by path identity. *Phys. Rev. Lett.* **118**, 080401 (2017).
35. Kysela, J., Erhard, M., Hochrainer, A., Krenn, M. & Zeilinger, A. Path identity as a source of high-dimensional entanglement. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 26118 (2020).

36. Nosrati, F., Castellini, A., Compagno, G. & Lo Franco, R. Robust entanglement preparation against noise by controlling spatial indistinguishability. *NPJ Quantum Inf.* **6**, 39 (2020).
37. Niemietz, D., Farrera, P., Langenfeld, S. & Rempe, G. Nondestructive detection of photonic qubits. *Nature* **591**, 570 (2021).
38. Johnson, B. R. *et al.* Quantum non-demolition detection of single microwave photons in a circuit. *Nat. Phys.* **6**, 663 (2010).
39. Munro, W. J., Nemoto, K., Beausoleil, R. G. & Spiller, T. P. High-efficiency quantum-nondemolition single-photon-number-resolving detector. *Phys. Rev. A* **71**, 033819 (2005).
40. Chin, S., Kim, Y.-S. & Lee, S. Graph picture of linear quantum networks and entanglement. *Quantum* **5**, 611 (2021).
41. Avron, J. E. & Kenneth, O. Entanglement and the geometry of two qubits. *Ann. Phys.* **324**, 470 (2009).

Acknowledgements

We thank Yong-Su Kim and Marek Żukowski for discussions and helpful comments. PB acknowledges support from the Polish-U.S. Fulbright Commission. MM acknowledges partial support from the Foundation for Polish Science (IRAP project, ICTQT, contract no. MAB/2018/5, co-financed by EU within Smart Growth Operational Programme).

Author contributions

All authors researched and wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to P.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Efficient linear-optical generation of a multipartite W state

Pawel Blasiak [1,2,*] Ewa Borsuk [1,†] Marcin Markiewicz [3,‡] and Yong-Su Kim [4,5,§]

[1]*Institute of Nuclear Physics, Polish Academy of Sciences, PL-31342 Kraków, Poland*

[2]*City, University of London, London EC1V 0HB, United Kingdom*

[3]*International Centre for Theory of Quantum Technologies, University of Gdańsk, PL-80308 Gdańsk, Poland*

[4]*Center for Quantum Information, Korea Institute of Science and Technology (KIST), Seoul 02792, Republic of Korea*

[5]*Division of Nano & Information Technology, KIST School, Korea University of Science and Technology, Seoul 02792, Republic of Korea*

(Received 2 March 2021; accepted 12 July 2021; published 3 August 2021)

A scheme is presented for generation of a multipartite W state for an arbitrary number of qubits. Based on a recent proposal of entanglement without touching, it serves to demonstrate the potential of particle indistinguishability as a useful resource of entanglement for practical applications. The devised scheme is efficient in design, meaning that it is built with linear optics without the need for auxiliary particles nor measurements. Yet, the success probability is shown to be highly competitive compared with the existing proposals (i.e., decreases polynomially with the number of qubits) and remains insensitive to particle statistics (i.e., has the same efficiency for bosons and fermions).

DOI: 10.1103/PhysRevA.104.023701

I. INTRODUCTION

Entanglement is a key quantum information resource which provides a basis for all modern developments in quantum technologies and remains a central theme in quantum foundations research [1–9]. Flexibility of entanglement generation is thus crucially important for practical implementations which require manipulation of an increasing number of qubits prepared in a desired state. Needless to say the challenges lie on both the experimental as well as the theoretical sides, with various factors deciding the usefulness of a given proposal. It is always a subtle interplay between the scaling properties for the efficiency of state generation and the resources measured by the complexity of the experimental design.

Here we focus on a prominent example of a multipartite entangled state, the W state, which for N qubits take the form,

$$W_N = \frac{1}{\sqrt{N}}\left(|\uparrow\downarrow\downarrow \ldots \downarrow\rangle + |\downarrow\uparrow\downarrow \ldots \downarrow\rangle + \ldots + |\downarrow \ldots \downarrow\downarrow\uparrow\rangle\right), \tag{1}$$

where $\{|\uparrow\rangle, |\downarrow\rangle\}$ is a computational basis. A remarkable property of this state is that the amount of entanglement shared between any of the qubits and all the rest is optimal, in a sense that entanglement is robust to the loss of one or more qubits in the system [10] (this should be compared with a multipartite GHZ state for which the loss of a single qubit forfeits any entanglement between all the rest). Notably, the W state belongs to a separate entanglement class under SLOCCs [11]. It also shows stronger nonlocal properties [12]. This makes the multipartite W states an interesting quantum resource, with a promise for practical applications in quantum information [13–15]. There have been a few implementations of the W state [16–20], but their complexity which grows with the number of qubits presents a significant challenge. For most theoretical proposals [21–23] the efficiency drops exponentially with N, with the recent exception in Ref. [24] (see also Ref. [25] for fermions).

In this paper, we develop a scheme for generation of the multipartite W state which can successfully compete with all known proposals, including the quantum erasure proposal in Ref. [24]. Our proposition is based on a *no-touching* paradigm for entanglement generation in linear optical circuits explicitly discussed in Ref. [26] (see also Refs. [27,28] for early indications of this idea and Refs. [29–31] for some particular realizations). This idea originates from the foundational question about the possibility of entanglement extraction from pure particle indistinguishability without any component from particle interactions. The problem has led to a special class of linear optical designs in which the particles traversing the circuit never touch one another over the entire evolution. Apparently, such experimental schemes provide an efficient platform for entanglement generation, which in the present paper is illustrated on the example of the multipartite W state. We also note that the no-touching designs are insensitive to particle statistics, that is, they are equally suitable for bosons and fermions.

The paper starts with some brief remarks on various factors relevant for the assessment of experimental proposals for state generation. Then we proceed with a detailed description of the no-touching scheme for generation of the multipartite W state for an arbitrary number of qubits. As the basis for our construction we use the dual-rail encoding of qubits and then give the corresponding design for polarization encoding. We also comment on experimental feasibility of the proposed schemes. The paper concludes with a discussion in which the

*pawel.blasiak@ifj.edu.pl
†ewa.borsuk@ifj.edu.pl
‡marcinm495@gmail.com
§yong-su.kim@kist.re.kr

efficiency of our design is compared with other proposals in the literature.

II. REMARKS ON EFFICIENCY COMPARISON

A standard tool for comparison of experimental designs is the efficiency denoted here by Eff_N, i.e., the success probability of obtaining the desired result (in our case it is a given state). The main factor reducing the efficiency comes form post-selection which occurs in many different guises in virtually every experimental design. A typical example are event-ready techniques for entanglement generation (e.g., entanglement swapping [32–34]) or more generally techniques based on coincidence counts (e.g., [35,36]). Then an evident figure of merit for multiqubit entanglement generation schemes is the scaling of the efficiency Eff_N with the increasing number of qubits N.

However, such an analysis blatantly ignores other details of the experimental design which may easily compromise the utility of a given proposal. A more thorough discussion should take into account the complexity as well as the resources required to implement a given proposal. Here we briefly point out some relevant factors in experimental designs which may have an impact that affects the overall assessment.

Suppose we consider a scheme that generates a given quantum state of N qubits which, for simplicity, are encoded in particle degrees of freedom (e.g., polarization, spin, or dual-rail encoding). It means that N particles carry the desired state ψ_{out} at the output. The efficiency Eff_N describes how often this state is produced, as a result of some well-defined post-selection procedure. Then, even if the complexity of linear operations in the circuit is not taken into account, the following features of the design still remain relevant:

(a) How many particles does the scheme start with? What input state ψ_{in} are they prepared in?

(b) Are intermediate measurements required in the design? Does the scheme involve feed-forward?

Re. (a). Every state generation scheme requires some initial multiparticle state ψ_{in} to start with, and some of them are easier to prepare than others. In particular, the fewer particles the better. Therefore every auxiliary particle in the input, on top of what is gained in the output, should count as an extra cost. Furthermore, if our goal is generation of entanglement, then any scheme starting with separable states ψ_{in}, like single-particle inputs $|1\rangle^{\otimes N}$, should be considered less demanding compared to those requiring entanglement to start with.

Re. (b). Measurements are a different kind of resource which introduces nonlinearity into the system. Thus the number of intermediate measurements adds to the overall cost of a given design. Even if the problems with detection inefficiency are ignored, the measurements introduce a stochastic element to the procedure. This either results in extra post-selection (cf. event-ready techniques [32,33]) or requires active correction using feed-forward. Needless to say that the latter presents a considerable technical challenge [34].

Clearly, the analysis solely based on the scaling of the efficiency Eff_N with the increasing number of particles N neglects all other relevant factors like those mentioned above. However, those other resources required in the design significantly contribute to the overall complexity (or cost) deciding the experimental utility of a given proposal. This makes it difficult to compare designs utilizing different resources, since in general it is not clear how to weigh between their costs. There is, however, one exception: From two designs having a comparable efficiency Eff_N, the one with no extra resources seems to be a better choice. This principle will allow one to appreciate the performance of the W state generation scheme constructed in the following section.

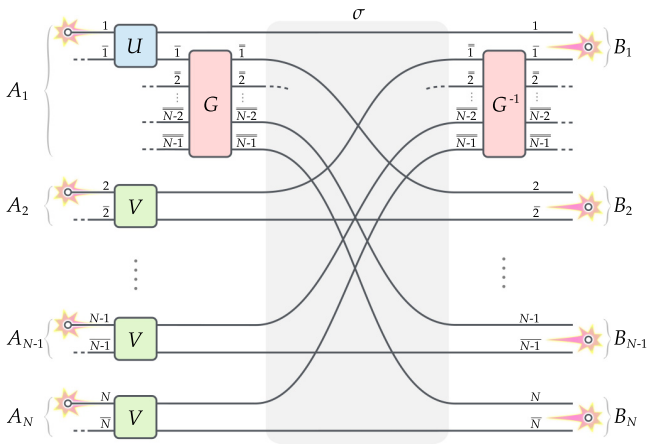


FIG. 1. **Generation of the W_N state in the no-touching scheme.** N independent particles injected into the circuit undergo a sequence of transformations: local unitaries $U \circ G, V, ..., V$ in the respective subsystems A_1,..., A_N followed by permutation of the paths σ and then local unitary G^{-1} in subsystem A_1. Post-selection on a single particle (coincidence) in each target qubit B_1,..., B_N generates the W_N state for the particular choice of unitaries in Eqs. (2) and (3) fulfilling condition Eq. (16).

III. NO-TOUCHING DESIGN FOR W STATE

A. Preliminaries

We follow the discussion of entanglement generation in the no-touching scheme proposed in Ref. [26]. Let us consider an optical scheme with $3N-2$ paths (or modes) grouped into N subsystems A_1,..., A_N with paths labeled as follows:

$$\begin{aligned}
&\text{System } A_1 = \{1, \overline{1}\} \cup \{\overline{\overline{2}}, \overline{\overline{3}}, ..., \overline{\overline{N-1}}\},\\
&\text{System } A_2 = \{2, \overline{2}\},\\
&\qquad\qquad ... \qquad\quad ...\\
&\text{System } A_N = \{N, \overline{N}\}.
\end{aligned}$$

See Fig. 1 for illustration. For better clarity we denote with a double bar the modes that will play an auxiliary role in subsystem A_1. Further, we assume that *target* qubits B_1,..., B_N are encoded in pairs of paths in the output (so called *dual-rail* qubits) for which we choose:

$$\text{Qubit } B_k = \{k, \overline{k}\} \qquad \text{for } k = 1, ..., N.$$

This means that the computational basis for qubit B_k is identified with a single particle in the respective path $|\uparrow\rangle_k \equiv a_k^\dagger|0\rangle$

and $|\downarrow\rangle_k \equiv a^\dagger_{\bar{k}}|0\rangle$, where $|0\rangle$ is the vacuum. Thus the general state of the qubit is encoded with a single particle in B_k as $\alpha\,|\uparrow\rangle_k + \beta|\downarrow\rangle_k \equiv (\alpha\, a^\dagger_k + \beta\, a^\dagger_{\bar{k}})|0\rangle$.

The protocol consists of the following stages (Fig. 1).

(1) Start with N independent particles with a single particle injected into each subsystem A_1,..., A_N.
(2) Local unitary on each subsystem A_1,..., A_N.
(3) Permutation of the paths σ,
(4) Local unitary in the output on subsystem A_1.
(5) Post-selection on a single particle (i.e., coincidence count) in the output qubits B_1,..., B_N.

Note that such a design guarantees that the particles do not *touch* one another over the entire evolution as discussed in Ref. [26]. For the discussion of its applicability to nonlocality tests see Ref. [37]. One particular property of such schemes is that their efficiency Eff_N is insensitive to particle statistics (i.e., the success probability is equal for bosons and fermions). In the following, we describe the no-touching protocol generating the W_N state Eq. (1) in the target qubits B_1,..., B_N.

B. Details of the protocol

Let us be precise about the no-touching protocol given in Fig. 1 and define three unitary transformations U, V, and G in the following form:

$$U = \begin{pmatrix} \alpha & \beta \\ \beta & -\alpha \end{pmatrix}, \qquad V = \begin{pmatrix} \delta & \varepsilon \\ \varepsilon & -\delta \end{pmatrix}, \tag{2}$$

and

$$G = \frac{1}{\sqrt{N-1}} \begin{pmatrix} 1 & \gamma_{1,2} & \ldots & \gamma_{1,N-1} \\ 1 & \gamma_{2,2} & \ldots & \gamma_{2,N-1} \\ \ldots & \ldots & \ldots & \ldots \\ 1 & \gamma_{N-1,2} & \ldots & \gamma_{N-1,N-1} \end{pmatrix}. \tag{3}$$

For simplicity we will assume that α, β, δ, and ε are real. Since U and V are unitary, we need to have $\beta^2 = 1-\alpha^2$ and $\varepsilon^2 = 1-\delta^2$. Numbers $\gamma_{k,l}$ are chosen arbitrarily so that G is unitary, and they will not play any role in our argument. Permutation of the modes depicted in Fig. 1 is defined as follows:

$$\sigma: \begin{cases} 1 \to 1, & \\ \overline{\overline{k}} \to k+1 & \text{for } k = 1, \ldots, N-1, \\ k \to \overline{\overline{k-1}} & \text{for } k = 2, \ldots, N, \\ \overline{k} \to \overline{k} & \text{for } k = 2, \ldots, N. \end{cases} \tag{4}$$

In this way we have specified all components of the protocol. For further convenience, let us write out all unitary transformations that will be used in our analysis:

$$a^\dagger_1 \xrightarrow{U} \alpha\, a^\dagger_1 + \beta\, a^\dagger_{\bar{1}}, \tag{5}$$

$$a^\dagger_k \xrightarrow{V} \delta\, a^\dagger_k + \varepsilon\, a^\dagger_{\bar{k}} \qquad \text{for } k = 2, 3, \ldots, N, \tag{6}$$

$$a^\dagger_{\bar{1}} \xrightarrow{G} \frac{1}{\sqrt{N-1}}\,(a^\dagger_1 + a^\dagger_2 + \ldots + a^\dagger_{\overline{N-1}}), \tag{7}$$

$$a^\dagger_{\bar{k}} \xrightarrow{G^{-1}} \frac{1}{\sqrt{N-1}}\,(a^\dagger_1 + \gamma^*_{k,2}\, a^\dagger_2 + \ldots + \gamma^*_{k,N-1}\, a^\dagger_{\overline{N-1}}) \qquad \text{for } k = 1, 3, \ldots, N-1. \tag{8}$$

Now, we are ready to trace the evolution of the input state of N independent particles injected into the circuit in Fig. 1 which ends with post-selection in the output. Postponing the questions of experimental difficulties to the next section, like the phase stability which should be maintained during the evolution, we may write

$$a^\dagger_1\, a^\dagger_2 \ldots a^\dagger_N\,|0\rangle \xrightarrow[\text{Eqs. (5)-(7)}]{U,G,V,\ldots,V} \left(\alpha\, a^\dagger_1 + \frac{\beta}{\sqrt{N-1}}\,(a^\dagger_{\bar{1}} + \ldots + a^\dagger_{\overline{N-1}})\right)(\delta\, a^\dagger_2 + \varepsilon\, a^\dagger_{\bar{2}}) \ldots (\delta\, a^\dagger_N + \varepsilon\, a^\dagger_{\bar{N}})|0\rangle \tag{9}$$

$$\xrightarrow[\text{Eq. (4)}]{\sigma} \left(\alpha\, a^\dagger_1 + \frac{\beta}{\sqrt{N-1}}\,(a^\dagger_2 + \ldots + a^\dagger_N)\right)(\delta\, a^\dagger_{\bar{1}} + \varepsilon\, a^\dagger_{\bar{2}}) \ldots (\delta\, a^\dagger_{\overline{N-1}} + \varepsilon\, a^\dagger_{\bar{N}})|0\rangle \tag{10}$$

$$\xrightarrow[\text{Eq. (8)}]{G^{-1}} \left(\alpha\, a^\dagger_1 + \frac{\beta}{\sqrt{N-1}}\,(a^\dagger_2 + \ldots + a^\dagger_N)\right) \tag{11}$$

$$\left(\frac{\delta}{\sqrt{N-1}}\left(a^\dagger_{\bar{1}} + \gamma^*_{1,2}\, a^\dagger_{\bar{2}} + \ldots + \gamma^*_{1,N-1}\, a^\dagger_{\overline{N-1}}\right) + \varepsilon\, a^\dagger_{\bar{2}}\right) \tag{12}$$

$$\ldots$$

$$\left(\frac{\delta}{\sqrt{N-1}}\,(a^\dagger_{\bar{1}} + \gamma^*_{N-1,2}\, a^\dagger_{\bar{2}} + \ldots + \gamma^*_{N-1,N-1}\, a^\dagger_{\overline{N-1}}) + \varepsilon\, a^\dagger_{\bar{N}}\right)|0\rangle \tag{13}$$

$$\xrightarrow{\text{post-selection}} \left(\alpha\varepsilon^{N-1}\, a^\dagger_1 a^\dagger_{\bar{2}} \ldots a^\dagger_{\bar{N}} + \frac{\beta\delta\varepsilon^{N-2}}{N-1} \sum_{k=2}^{N} a^\dagger_k\, a^\dagger_{\bar{2}} \ldots a^\dagger_{\overline{k-1}}\, a^\dagger_{\bar{1}}\, a^\dagger_{\overline{k+1}} \ldots a^\dagger_{\bar{N}}\right)|0\rangle \tag{14}$$

$$\overset{Eq.(16)}{=} \sqrt{\frac{\delta^2\,(1-\delta^2)^{N-1}}{\delta^2 + (N-1)^2\,(1-\delta^2)}}\,(|\uparrow\downarrow\ldots\downarrow\rangle + |\downarrow\uparrow\downarrow\ldots\downarrow\rangle + \ldots + |\downarrow\ldots\downarrow\uparrow\downarrow\rangle + |\downarrow\ldots\downarrow\downarrow\uparrow\rangle). \tag{15}$$

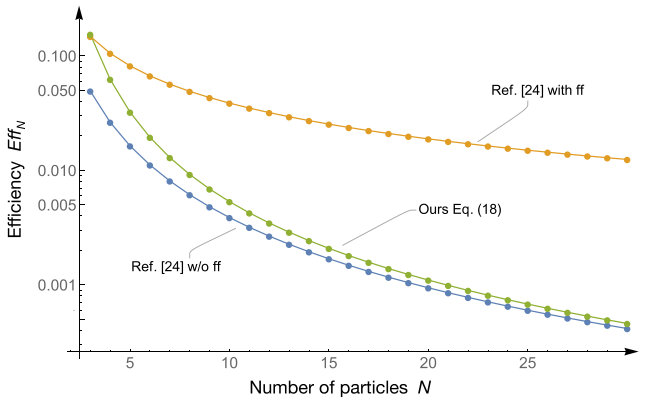


FIG. 2. **Optimal efficiency of the scheme.** The success probability for generation of the W_N state in our scheme Eq. (18) compared with the scheme with auxiliary particle and quantum erasure described in Ref. [24] [without and with active feed-forward; see Eqs. (12) and Eq. (14) therein].

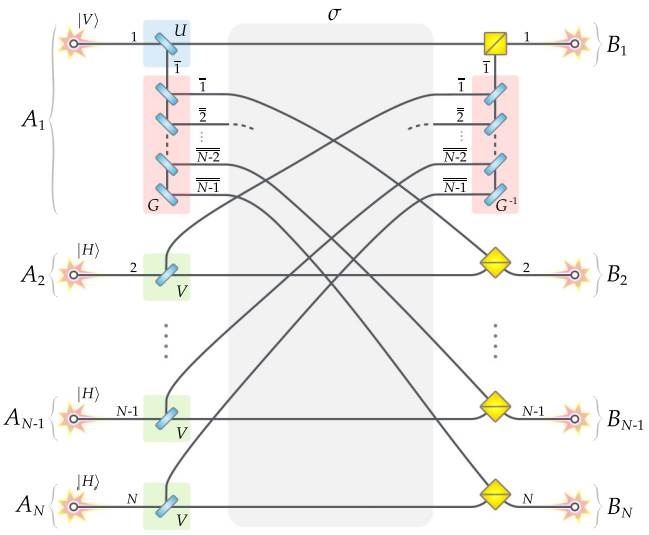


FIG. 3. **Optical scheme with photon polarization.** Rewriting of the proposal in Fig. 1 into photon polarization qubits. In that case, each pair of paths $\{k, \bar{k}\}$ in dual-rail encoding gets replaced with two polarization modes $\{|H\rangle_k, |V\rangle_k\}$, with spatial merging obtained by polarizing beam splitters (PBS) in the output. Then the unitaries U, V, G, and G^{-1} are implemented by appropriate choice of beam splitters (BS). The analysis of this circuit parallels the dual-rail case in Fig. 1 and the resulting efficiencies are clearly the same.

Note that the product structure in Eqs. (11)–(13) allows for quick identification of terms that remain after post-selection. Since we are interested only in cases with a single particle in each subsystem $B_k = \{k, \bar{k}\}$, then for each creator in the first bracket Eq. (11) there is only one choice of creators in the remaining brackets Eq. (12) and (13) which fulfills the post-selection condition: a single particle (creator) in each subsystem B_k (notice that the modes with double bar can be dropped altogether). What is left are only the terms in Eq. (14). The last equality in Eq. (15) holds for bosons when all coefficients are equal,

$$\alpha\varepsilon^{N-1} = \frac{\beta\delta\varepsilon^{N-2}}{N-1}, \tag{16}$$

which holds for the choice $\alpha^2 = \frac{\delta^2}{\delta^2+(N-1)^2(1-\delta^2)}$. Thus we obtain the W_N state Eq. (1) in the output. We remark that for fermions all terms except the first one in Eq. (15) get the "−" sign which can be easily corrected by phase shift $e^{i\pi}$ in the first path (either in the input or output).

Note that the expression in Eq. (15) is unnormalized due to post-selection which projects on the subspace with a single particle in each channel $B_1, \ldots, B_N$. From the normalization we can read out the success probability (efficiency) of the process which is equal to

$$\mathrm{Eff}_N(\delta) = \frac{N\,\delta^2\,(1-\delta^2)^{N-1}}{\delta^2 + (N-1)^2\,(1-\delta^2)}. \tag{17}$$

Since there is one free parameter left in the protocol we can optimize over δ and obtain the maximal efficiency:

$$\mathrm{Eff}_N = \max_\delta \mathrm{Eff}_N(\delta) \sim \frac{e^{-1}}{N^2} + \frac{7\,e^{-1}}{2\,N^3} + o\left(\frac{1}{N^4}\right). \tag{18}$$

The result is plotted in Fig. 2 and compared with the efficiency of the quantum erasure scheme in Ref. [24]. See Appendix for explicit calculation. Let us note that the obtained efficiency is insensitive to particle statistics, which is the same for bosons and fermions. This follows from two observations: (i) in the passive linear optics regime the unitary evolution is governed by the same set of equations Eqs. (5)–(8) regardless of the statistics [38], and (ii) in the post-selected sector the same terms survive for both types of particles (due to the no-touching feature of the scheme [26]).

We remark that the above discussion based on dual-rail encoding of qubits gives a generic pattern which directly translates into other realizations. For illustration, in Fig. 3 we give analogous design for photon polarization qubits using the scheme in Fig. 1 as a template.

C. Experimental feasibility

Our N particle W state generation protocol requires N identical single photons and the phase stabilized linear optical networks. We remark that such technical demands can be fulfilled with the current quantum photonics technology. For example, it has been shown that tens of identical single photons can be generated either via spontaneous parametric down-conversion (SPDC) [35] or from a quantum dot [39]. The temporal multiplexing technique with SPDC also provides a promising avenue to scale up the number of identical single photon generation [40]. The phase stabilized linear optical networks can be achieved with the active feedback control. For instance, a huge phase stabilized interferometer with an arm length of a few hundreds of kilometers has been implemented in the context of twin-field quantum key distribution [41]. We also note that the integrated quantum photonics can provide an efficient way to implement complicated linear optical networks with the excellent phase stability [9]. Therefore, our cost efficient W state generation protocol can be implemented with tens of photons with the current quantum photonics technology.

IV. DISCUSSION

The presented protocol for generation of the W state is cost effective. It only requires linear optics and post-selection to get the N particle W state from N independent particles in the input. The efficiency of state generation scales polynomially $\mathrm{Eff}_N \sim 1/N^2$ with the number of particles. This leaves behind most proposals in the literature which use quantum fusion techniques [16–20] for the iterative construction of the larger W_N states from smaller ones $W_{M<N}$, which inherently suffer from the exponential decrease of efficiency $\mathcal{O}^{-N}$. Our proposal clearly benefits from the direct construction of the state rather than the use of iterative techniques. It is also more economical with the respect of resources required for the experimental implementation.

Interestingly, there is a recent proposal of multipartite W state generation based on quantum erasure [24] (see also Ref. [25] for fermions). It is to our knowledge the most efficient protocol whose efficiency scales polynomially like $1/N^2$, with a potential for further improvement to $1/N$ by feed-forward with active state correction on each qubit (see Fig 2). Note, however, that this scheme requires one auxiliary particle in the input on which appropriate measurement is made, i.e., $N+1$ particles are needed to produce the N particle W_N state. Clearly, this is an additional cost both in terms of particles and measurements that a fair comparison should take into account. Neither of those are required in our design and yet it performs slightly better (cf. Appendix). As noted, even if post-processing of each qubit based on active feed-forward in the quantum erasure scheme in Ref. [24] increases the efficiency, this comes with a considerable extra cost which scales linearly with the number of particles N (and thus compromising the potential gain). Needless to say that active feed-forward poses a nontrivial experimental challenge.

Finally, we remark that the presented approach is insensitive to particle statistics, i.e., works with equal efficiency for fermions, bosons, or anyons. This is a generic property of the no-touching designs [26]. Moreover, it can be safely used for Bell nonlocality tests, since the post-selection scheme follows the all-but-one principle discussed in Ref. [37], which assures no post-selection loophole. Therefore our protocol can be safely utilized as a part of device-independent protocols in quantum technological applications.

ACKNOWLEDGMENTS

We thank M. Żukowski for discussions and helpful comments. P.B. acknowledges support from the Polish National Agency for Academic Exchange in the Bekker Scholarship Programme. We acknowledge partial support by the Foundation for Polish Science (IRAP project, ICTQT, contract No. MAB/2018/5, co-financed by EU within Smart Growth Operational Programme). Y.-S.K. acknowledges support from the National Research Foundation of Korea (Grant No. 2019M3E4A1079777) and the KIST ORP program (2E31021).

APPENDIX: MAXIMAL EFFICIENCY Eff_N IN EQ. (18)

Finding the maximum in Eq. (17) is straightforward. It boils down to differentiating and choosing the solution such that $|\delta| \leqslant 1$. Thus we get

$$\frac{\partial \mathrm{Eff}_N(\delta)}{\partial \delta} = 0 \Rightarrow \delta^2_{\max} = \frac{1 - N + \sqrt{\frac{N^3 - 6N^2 + 13N - 8}{N}}}{4 - 2N}, \tag{A1}$$

and hence Eq. (18) takes the form,

$$\mathrm{Eff}_N = \mathrm{Eff}_N(\delta_{\max}) = \frac{N\,\delta^2_{\max}\left(1-\delta^2_{\max}\right)^{N-1}}{\delta^2_{\max} + (N-1)^2\left(1-\delta^2_{\max}\right)}. \tag{A2}$$

This result is plotted in Fig. 2.

For completeness, we also give asymptotic expansion of the above expression which reads

$$\mathrm{Eff}_N \sim \frac{e^{-1}}{N^2} + \frac{7\,e^{-1}}{2\,N^3} + o\left(\frac{1}{N^4}\right). \tag{A3}$$

This should be compared with the corresponding efficiency for the recent proposal with quantum erasure in Ref. [5] (to our knowledge the most efficient proposal for W state generation in the literature). Equation (14) therein gives

$$\mathrm{Eff}_N^{(\mathrm{Kim}\,et\,al.)} \sim \frac{e^{-1}}{N^2} + \frac{e^{-1}}{2\,N^3} + o\left(\frac{1}{N^4}\right). \tag{A4}$$

We note that although the use of active feed-forward in the scheme of Kim *et al.* [24] increases the efficiency by factor N, it comes at an incomparable cost which scales with the number of particles N, too.

[1] J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics* (Cambridge University Press, Cambridge, 1987).
[2] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).
[3] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Quantum entanglement, Rev. Mod. Phys. **81**, 865 (2009).
[4] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Bell nonlocality, Rev. Mod. Phys. **86**, 419 (2014).
[5] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, Quantum cryptography, Rev. Mod. Phys. **74**, 145 (2002).
[6] T. D. Ladd, F. Jelezko, R. Laflamme, Y. Nakamura, C. Monroe, and J. L. O'Brien, Quantum computers, Nature (London) **464**, 45 (2010).
[7] A. Aspuru-Guzik and P. Walther, Photonic quantum simulators, Nat. Phys. **8**, 285 (2012).
[8] J.-W. Pan, Z.-B. Chen, C.-Y. Lu, H. Weinfurter, A. Zeilinger, and M. Żukowski, Multiphoton entanglement and interferometry, Rev. Mod. Phys. **84**, 777 (2012).
[9] J. Wang, F. Sciarrino, A. Laing, and M. G. Thompson, Integrated photonic quantum technologies, Nat. Photonics **14**, 273 (2020).

[10] M. Koashi, V. Bužek, and N. Imoto, Entangled webs: Tight bound for symmetric sharing of entanglement, Phys. Rev. A **62**, 050302(R) (2000).

[11] W. Dür, G. Vidal, and J. I. Cirac, Three qubits can be entangled in two inequivalent ways, Phys. Rev. A **62**, 062314 (2000).

[12] A. Sen(De), U. Sen, M. Wieśniak, D. Kaszlikowski, and M. Żukowski, Multiqubit W states lead to stronger nonclassicality than Greenberger-Horne-Zeilinger states, Phys. Rev. A **68**, 062306 (2003).

[13] P. Agrawal and A. Pati, Perfect teleportation and superdense coding with *W* states, Phys. Rev. A **74**, 062320 (2006).

[14] C. Zhu, F. Xu, and C. Pei, W-state analyzer and multi-party measurement-device-independent quantum key distribution, Sci. Rep. **5**, 17449 (2015).

[15] V. Lipinska, G. Murta, and S. Wehner, Anonymous transmission in a noisy quantum network using the *W* state, Phys. Rev. A **98**, 052320 (2018).

[16] M. Eibl, N. Kiesel, M. Bourennane, C. Kurtsiefer, and H. Weinfurter, Experimental Realization of a Three-Qubit Entangled W State, Phys. Rev. Lett. **92**, 077901 (2004).

[17] H. Mikami, Y. Li, K. Fukuoka, and T. Kobayashi, New High-Efficiency Source of a Three-Photon W State and its Full Characterization Using Quantum State Tomography, Phys. Rev. Lett. **95**, 150404 (2005).

[18] T. Tashima, T. Wakatsuki, S. K. Özdemir, T. Yamamoto, M. Koashi, and N. Imoto, Local Transformation of Two Einstein-Podolsky-Rosen Photon Pairs into a Three-Photon W State, Phys. Rev. Lett. **102**, 130502 (2009).

[19] T. Tashima, T. Kitano, S. K. Özdemir, T. Yamamoto, M. Koashi, and N. Imoto, Demonstration of Local Expansion Toward Large-Scale Entangled Webs, Phys. Rev. Lett. **105**, 210503 (2010).

[20] B. Fang, M. Menotti, M. Liscidini, J. E. Sipe, and V. O. Lorenz, Three-Photon Discrete-Energy-Entangled W State in an Optical Fiber, Phys. Rev. Lett. **123**, 070508 (2019).

[21] X.-P. Zang, M. Yang, F. Ozaydin, W. Song, and Z.-L. Cao, Deterministic generation of large scale atomic W states, Opt. Express **24**, 12293 (2016).

[22] N.-N. Wu and M. Jiang, A highly efficient scheme for joint remote preparation of multi-qubit W state with minimum quantum resource, Quantum Inf. Process. **17**, 340 (2018).

[23] S. Bugu, F. Ozaydin, T. Ferrus, and T. Kodera, Preparing multipartite entangled spin qubits via pauli spin blockade, Sci. Rep. **10**, 3481 (2020).

[24] Y.-S. Kim, Y.-W. Cho, H.-T. Lim, and S.-W. Han, Efficient linear optical generation of a multipartite W state via a quantum eraser, Phys. Rev. A **101**, 022337 (2020).

[25] B. Bellomo, R. Lo Franco, and G. Compagno, N identical particles and one particle to entangle them all, Phys. Rev. A **96**, 022319 (2017).

[26] P. Blasiak and M. Markiewicz, Entangling three qubits without ever touching, Sci. Rep. **9**, 20131 (2019).

[27] B. Yurke and D. Stoler, Bell's-inequality experiments using independent-particle sources, Phys. Rev. A **46**, 2229 (1992).

[28] B. Yurke and D. Stoler, Einstein-Podolsky-Rosen Effects from Independent Particle Sources, Phys. Rev. Lett. **68**, 1251 (1992).

[29] I. Neder, N. Ofek, Y. Chung, M. Heiblum, D. Mahalu, and V. Umansky, Interference between two indistinguishable electrons from independent sources, Nature (London) **448**, 333 (2007).

[30] Y.-S. Kim, T. Pramanik, Y.-W. Cho, M. Yang, S.-W. Han, S.-Y. Lee, M.-S. Kang, and S. Moon, Informationally symmetrical Bell state preparation and measurement, Opt. Express **26**, 29539 (2018).

[31] L. Ju, M. Yang, N. Paunković, W.-J. Chu, and Z.-L. Cao, Creating photonic GHZ and W states via quantum walk, Quantum Inf. Process. **18**, 176 (2019).

[32] M. Żukowski, A. Zeilinger, M. A. Horne, and A. K. Ekert, "Event-Ready-Detectors" Bell Experiment via Entanglement Swapping, Phys. Rev. Lett. **71**, 4287 (1993).

[33] J.-W. Pan, D. Bouwmeester, H. Weinfurter, and A. Zeilinger, Experimental Entanglement Swapping: Entangling Photons That Never Interacted, Phys. Rev. Lett. **80**, 3891 (1998).

[34] X.-S. Ma, T. Herbst, Thomas Scheidl, D. Wang, S. Kropatschek, W. Naylor, B. Wittmann, A. Mech, J. Kofler, E. Anisimova, V. Makarov, T. Jennewein, R. Ursin, and A. Zeilinger, Quantum teleportation over 143 kilometres using active feed-forward, Nature (London) **489**, 269 (2012).

[35] H.-S. Zhong, Y. Li, W. Li, L.-C. Peng, Z.-E. Su, Y. Hu, Y.-M. He, X. Ding, W. Zhang, H. Li, L. Zhang, Z. Wang, L. You, X.-L. Wang, X. Jiang, L. Li, Y.-A. Chen, N.-L. Liu, C.-Y. Lu, and J.-W. Pan, 12-Photon entanglement and Scalable Scattershot Boson Sampling with Optimal Entangled-Photon Pairs from Parametric Down-Conversion, Phys. Rev. Lett. **121**, 250505 (2018).

[36] M. Krenn, A. Hochrainer, M. Lahiri, and A. Zeilinger, Entanglement by Path Identity, Phys. Rev. Lett. **118**, 080401 (2017).

[37] P. Blasiak, E. Borsuk, and M. Markiewicz, On safe post-selection for Bell nonlocality: Causal diagram approach, arXiv:2012.7285 (2021).

[38] M. C. Tichy, Interference of identical particles from entanglement to boson-sampling, J. Phys. B: At. Mol. Opt. Phys. **47**, 103001 (2014).

[39] H. Wang, J. Qin, X. Ding, M.-C. Chen, S. Chen, X. You, Y.-M. He, X. Jiang, L. You, Z. Wang, C. Schneider, J. J. Renema, S. Hofling, C.-Y. Lu, and J.-W. Pan, Boson Sampling with 20 Input Photons and a 60-Mode Interferometer in a 10^{14}-Dimensional Hilbert Space Demonstration, Phys. Rev. Lett. **123**, 250503 (2019).

[40] F. Kaneda and P. G. Kwiat, High-efficiency single-photon generation via large-scale active time multiplexing, Sci. Adv. **5**, eaaw8586 (2019).

[41] S. Wang, D.-Y. He, Z.-Q. Yin, F.-Y. Lu, C.-H. Cui, W. Chen, Z. Zhou, G.-C. Guo, and Z.-F. Han, Beating the Fundamental Rate-Distance Limit in a Proof-Of-Principle Quantum Key Distribution System, Phys. Rev. X **9**, 021046 (2019).

Causal reappraisal of the quantum three-box paradox

Pawel Blasiak[*] and Ewa Borsuk[†]
Institute of Nuclear Physics Polish Academy of Sciences, PL-31342 Kraków, Poland

(Received 29 July 2021; accepted 23 November 2021; published 10 January 2022)

The quantum three-box paradox is a prototypical example of some bizarre predictions for intermediate measurements made on *pre- and postselected* systems. Although in principle those effects can be explained by measurement disturbance, it is not clear what mechanisms are required to fully account for the observed correlations. In this paper, this paradox is scrutinized from the causal point of view. We consider an array of potential causal structures behind the experiment, eliminating those without enough explanatory power. A distinction is made between the propagation in the system of just the *measurement outcome* and the information about the *full measurement context*. We also discuss the consequences of the *realist* position in which preexisting values are revealed by measurements. Interestingly, the answers depend crucially on whether the original version of the paradox is considered or its extension where the third box is allowed for inspection too. This illustrates the richness of the paradox which is better appreciated from the causal perspective.

DOI: 10.1103/PhysRevA.105.012207

I. INTRODUCTION

A paradox builds upon a conflict between observed facts and preconceptions that we hold about them, with a view to elicit a revision of the latter for a deeper understanding of a given problem or phenomenon. On the one hand, this may be just a warning about a superficial understanding of the mathematics, which if correctly applied does not lead to any contradictions. It is particularly true about the problems involving probability. On the other hand, a paradox can be an indication of a deeper misconception regarding the mechanisms at work, and thus potentially revealing something new about the nature itself. Therein lies the interest for the foundational issues of quantum theory and, in particular, the question about the causal nature behind its predictions. Notably, some remarkable results on quantum nonlocality [1–3] or contextuality [4–6] are prime examples with a paradox and causality at the background.

A *three-box paradox* [7] is a flagship example of the *pre- and postselection* (PPS) scenarios, in which some surprising predictions about intermediate measurements are made. It was originally proposed as an illustration of the so-called ABL rule [8] (after Aharonov, Bergmann, and Lebowitz). For the three-box paradox case it makes a strange prediction regarding the position of a particle which is always found where it is looked for. This has sparked controversy regarding the nature of the paradox and conclusions that can be drawn from this bizarre effect [9–15]. The first objection concerns the presence of postselection in the experiment, since the rejection of data is a potential source of noncausal correlations known as a *selection bias* [16]. The second problem stems from the possible role of *measurement disturbance* in the experiment, since in this case the disturbance can propagate in the system making the information about the intermediate measurements available at the moment of postselection. Those issues certainly affect interpretation of the paradox and thus need careful reassessment within a proper conceptual framework.

In this paper we tread the path eloquently expressed in Pearl's [17] conjecture that "*human intuition is organized around casual, not statistical, relations*." It suggests that in an attempt at resolving a paradox one should rather focus on causal mechanisms behind the observed correlations. Not only does this give a way to the bottom of the paradox by explicating the implicit assumptions that we make, but it sometimes may even offer something new about the causal mechanisms at work. Notably, the causal approach has recently gained a solid mathematical foundation in the works of Pearl and others [16,18,19] (which goes along similar lines as introduced by Bell [1]). Some remarkable results in the field of causal inference pertinent to the present paper include *d-separation rules* and *instrumental inequalities*. This novel approach has helped in resolving various conundrums in observational studies in epidemiology, computer science, and social sciences [20–22]. Despite a fairly recent development, mostly outside of physics, those methods have already influenced the research in quantum foundations (see, e.g., Refs. [23–31]).

In this paper, we employ the tools of causal inference to analyze measurement disturbance and its impact on postselection in the three-box paradox. This approach allows differentiating between the various mechanisms in which measurement disturbance can propagate. We also bring to light some implicit assumptions about realism that are typically made, which explains where the clash with our intuition might come from.

II. THREE-BOX PARADOX IN A NUTSHELL

Consider an experiment with a system prepared at time t_0 in some state initial state ρ_0 on which at a later time t_2 a projective measurement $\mathcal{M}_2$ is made checking for state Π^{post}. Let us agree to retain only the positive results $M_2 = 1$, which are deemed a success, and discard all the rest $M_2 = 0$. This is the so-called PPS scenario. To make it more interesting we

[*]pawel.blasiak@ifj.edu.pl
[†]ewa.borsuk@ifj.edu.pl

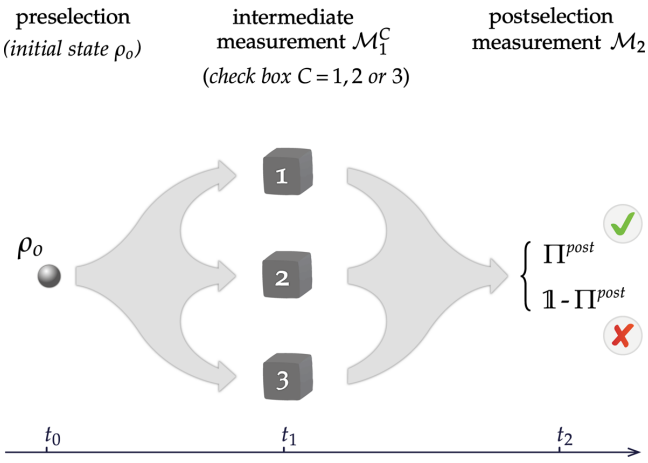


FIG. 1. Three-box experiment. Consider a particle that can be in one of three boxes labeled 1, 2, and 3. The system is preselected in state ρ_0 and postselected in state Π^{post}. In each experimental trial we choose a box $C = 1, 2, 3$ checking whether the particle is there or not. Quantum mechanics makes a puzzling prediction that whichever box $C = 1$ or 2 we choose to look at, the particle will be always found there [see Eq. (4)].

allow ourselves to make a measurement $\mathcal{M}_1^c$ in some intermediate time t_1 ($t_0 < t_1 < t_2$) described by a projection-valued measure (PVM) $\{\Pi_i^c\}$, where C stands for the *choice* of measurement in a given experimental run. Then, the conditional probability of obtaining outcome $M_1 = i$ in the PPS scenario is given by the ABL rule [8]

$$P(M_1 = i|M_2 = 1, C) = \frac{\mathrm{Tr}[\Pi^{\text{post}}\,\Pi_i^C\,\rho_0\,\Pi_i^C]}{\sum_l \mathrm{Tr}[\Pi^{\text{post}}\,\Pi_l^C\,\rho_0\,\Pi_l^C]}. \tag{1}$$

It is a straightforward application of Bayes' theorem to the joint probability

$$P(M_1 = i, M_2 = j|C) = \mathrm{Tr}\big[\Pi_j^{\text{post}}\,\Pi_i^c\,\rho_0\,\Pi_i^c\big], \tag{2}$$

which is obtained by the usual von Neumann–Lüders rule. Here the final measurement $\mathcal{M}_2$ is described by the PVM $\{\Pi_j^{\text{post}}\}_{j=0,1} \equiv \{\mathbb{1} - \Pi^{\text{post}}, \Pi^{\text{post}}\}$.

A three-box paradox [7] is a specific realization of the PPS scenario in which the intermediate measurements $\mathcal{M}_1^c$ are assigned deterministic conditional outcomes. It concerns a single particle that can be localized in one of three boxes labeled 1, 2, and 3 described by the respective quantum states $|1\rangle$, $|2\rangle$, and $|3\rangle$. Suppose in the intermediate measurements we check whether the particle is in a given box $C = 1, 2, 3$, or not, which is implemented by the PVM $\{\Pi_i^c\}_{i=0,1} \equiv \{\mathbb{1} - |C\rangle\langle C|, |C\rangle\langle C|\}$. See Fig. 1. Now, if we choose for the pre- and postselected states, $\rho_0 = |\phi\rangle\langle\phi|$ and $\Pi^{\text{post}} = |\psi\rangle\langle\psi|$, the following nonorthogonal pair,

$$|\phi\rangle = \tfrac{|1\rangle+|2\rangle+|3\rangle}{\sqrt{3}} \quad \text{and} \quad |\psi\rangle = \tfrac{|1\rangle+|2\rangle-|3\rangle}{\sqrt{3}}, \tag{3}$$

then, from the ABL rule Eq. (1), we get

$$P(M_1 = 1|M_2 = 1, C) = 1 \qquad \text{for } C = 1, 2. \tag{4}$$

Hence, we have a paradoxical conclusion that whatever box we check, $C = 1$ or 2, the particle is always there.

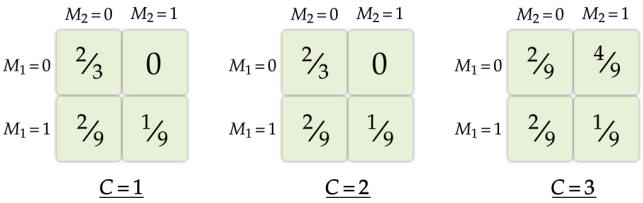


FIG. 2. Full statistics in three-box experiment. The joint probability $P(M_1, M_2|C)$ of obtaining measurement outcomes $M_1 = 0, 1$ (i.e., particle not found or found) and $M_2 = 0, 1$ (i.e., postselection is a failure or success) for the choice of experiment $C = 1, 2, 3$ (i.e., which box to check).

The full statistics observed in the experiment is given in Fig. 2, which readily follows from Eq. (2). This shows that for $C = 1, 2$ postselection succeeds with the probability equal to $P(M_2 = 1|C) = 1/9$.[1] Let us note in advance that although the original formulation of the paradox in Eq. (4) concerns just two (out of three possible) experimental choices $C = 1, 2$, it becomes more revealing and weird when we look at the full statistics $C = 1, 2, 3$.

III. CAUSAL PICTURE OF THE THREE-BOX EXPERIMENT

Let us consider the possible causal structures hiding behind the experiment which will be further assessed against their capacity for generating the three-box statistics in Fig. 2. We need to decide about the variables deemed relevant for the description of the experiment and then ponder their causal relationships.

A. Pure causal setting

In the description of the three-box experiment there are three *observed* variables:

C : choice of measurement setting ($C = 1, 2, 3$),

M_1 : outcome of measurement $\mathcal{M}_1^c$ ($M_1 = 0, 1$),

M_2 : outcome of measurement $\mathcal{M}_2$ ($M_2 = 0, 1$).

Furthermore, let us postulate the existence of some *unobserved* variable:

Λ : hidden (or latent) variable.

This variable is aimed to describe any other factors relevant for the experiment (e.g., the details of the preparation procedure). It is left unspecified in order not to restrict the range of possible explanations behind the observed correlations. We call it a *pure causal* setting as it involves the least number of assumptions (this should be compared with the more restricted framework below).

B. Realist causal setting

It is very instructive to consider in parallel the additional *realism* assumption. This is a position which derives from the worldview in which physical objects and their properties exist,

[1] We remark that this probability is the same as when no intermediate measurement $\mathcal{M}_1^c$ is made, in which case we also have $P(M_2 = 1|$ without $\mathcal{M}_1^c) = \mathrm{Tr}[\Pi^{\text{post}}\,\rho_0] = |\langle\psi\,|\,\phi\rangle|^2 = 1/9$.

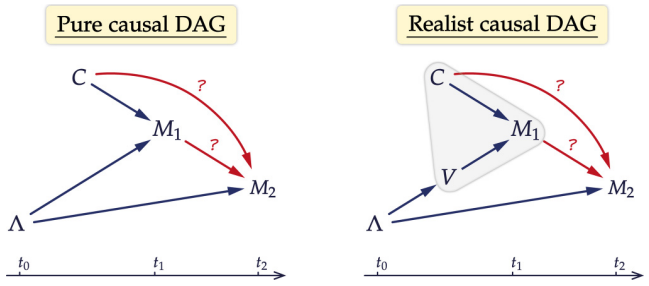


FIG. 3. Causal diagrams for the three-box experiment. In both diagrams, the variables C, M_1, and M_2 are observed in the experiment, whereas Λ and V are unobserved (latent or hidden) variables. These are directed acyclic graphs (DAGs) which are allowed by the temporal structure of the experiment (no retro-causation) and assuming C to be a free variable. The diagram on the right includes an additional structure (shaded in gray) which reflects the realism assumption. In both diagrams the red arrows (with question marks) are responsible for different types of measurement disturbance propagating in the system, i.e., whether it is just the outcome M_1 or the full measurement context C that affects M_2.

and the measurements reveal those preexisting values. In our case this view boils down to the existence of an additional variable

$$V : \text{position of the particle } (V = 1, 2, 3),$$

having a definite value before the measurement is made (possibly a derivative of the hidden variable Λ). Since the measurement $\mathcal{M}_1^c$ answers the question "Is the particle in a given box C?", we have the following consistency condition:

$$M_1(C, V) := \delta_{C,V}. \tag{5}$$

It is a direct expression of the requirement that the measurement reveals the property of the particle being in a given box. We remark that those additional structural components make the *realist causal* setting more restrictive compared to the *pure causal* setting, as we shall see shortly.

C. Causal diagrams for the experiment

Let us draw the diagrams compatible with the above two descriptions. In Fig. 3 the arrows represent cause-and-effect relationships between the variables. Observe that the temporal structure of the experiment allows us to eliminate certain arrows in the diagrams. Namely, we assume only *forward-in-time causation*. Also, since the choice of measurement is considered to be a *free variable*, we assume there is no arrow incoming to C. [Note that, although in principle in the *realist causal directed acyclic graph* (DAG) (right) there could be an arrow $V \to M_2$, it has not been drawn since we can always incorporate it in the arrow $\Lambda \to M_2$ (by appropriately modifying Λ).]

We note that the diagrams in Fig. 3 include *all* arrows compatible with the experiment. In this paper we pose the question about the *necessity* of arrows $M_1 \to M_2$ and $C \to M_2$. Both are responsible for the causal effects of the intermediate measurement $\mathcal{M}_1^c$ in the experiment. The lack of both arrows means no measurement disturbance. Conversely, their presence is a sign of different types of disturbance propagating in the system, i.e., whether it is just the measurement outcome M_1 or the full context specified by the choice of measurement (parameter) C that affects the final measurement M_2. Using the terminology borrowed from the analysis of Bell nonlocality [32], we call those arrows as

$$M_1 \to M_2 \ : \text{outcome dependence},$$
$$C \to M_2 \ : \text{parameter dependence}$$

Having specified causal structures of interest we may assess their potential for explaining the statistics of the three-box paradox in Fig. 2.

IV. MAIN RESULTS

A. Inspection of causal structures

In the causal inference field we are interested in verifying whether a given causal structure can explain the observed experimental behavior, i.e., whether the statistics can be reproduced by some *structural causal model* (SCM) consistent with a given *causal DAG* (see Refs. [16,18–29,31]). Since adding arrows to the diagram extends its expressive power, we are looking for structures with the fewest number of arrows which can still explain the observed experimental behavior.

As noted, the realist causal setting is more restrictive than the pure causal setting (see Fig. 3), i.e., not all behaviors compatible with the pure causal DAG (left) are admitted by the realist causal DAG (right). In the following we consider both cases separately. Furthermore, we also distinguish between the statistics in the three-box experiment in Fig. 2 for the *full* choice of three measurements $C = 1, 2, 3$, and the case *limited* to the two choices $C = 1, 2$ (as in the original exposition of the paradox). This will make an interesting case regarding our perception of the paradox and its further ramifications as explained below.

Our results are summarized in Fig. 4. For the proofs of necessity of the respective arrows see Appendix A (which employs the tools of causal inference [16]: the d-separation criterion and instrumental inequalities). The proofs of sufficiency are given in Appendix B (this requires explicit construction of SCMs in each case).

B. Conclusions

Having analyzed possible causal explanations of the three-box experiment it is natural to ask why people find it surprising. Notably, the original formulation of the paradox [7] concerns just two (out of three possible) experimental choices $C = 1, 2$; here the postselected behavior is deterministic, which makes it better suited for human judgment. In this case the paradox seems to arise from the tension between the assumption of realism and the deceptive impression, from how the paradox is phrased, as to the lack of measurement disturbance in the experiment. We showed that both assertions are intrinsically contradictory. The requirement of realism necessitates measurement disturbance of some sort [see Fig. 4(b)], i.e.,

$$\text{Realism} \underset{C=1,2}{\Rightarrow} \text{measurement disturbance}.$$

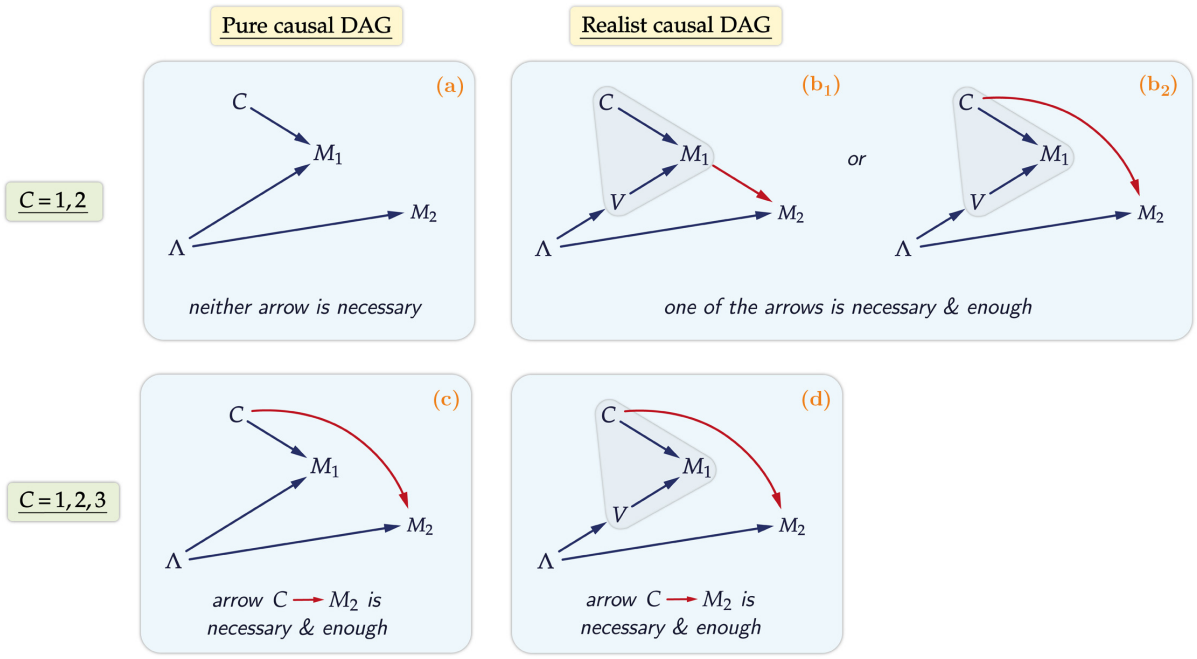


FIG. 4. Summary of the results. The table answers the question of which red arrows in the two causal diagrams in Fig. 3 can be removed while still retaining their capacity for generating the statistics in the three-box experiment in Fig. 2. There is a difference whether the full statistics is considered $C = 1, 2, 3$ or just its part related to the "paradoxical" choice of measurements $C = 1, 2$. This shows what kind of measurement disturbance (*outcome* vs *parameter dependence*) is required depending on the preferred worldview (pure vs realist) and the selection of measurements under consideration.

Accordingly, if the disturbance is taken on board, then the paradoxical correlations in the postselected regime can be explained as an instance of the *selection bias* [16] [see the proof of Fig. 4(b) in Appendix A].

Interestingly, in the pure causal setting (no realism) the paradox can be explained without measurement disturbance of any sort. In this case we showed, by constructing the explicit SCM, that confounding is just enough to explain the effect [see Fig. 4(a)], i.e.,

$$\text{No realism} \underset{C=1,2}{\not\Rightarrow} \text{measurement disturbance}.$$

Thus without claiming realism the question of measurement disturbance turns out immaterial, and hence the paradox becomes less of a surprise.

As noted, both conclusions above concern the original formulation of the paradox with just two boxes being considered for inspection ($C = 1, 2$).

Our discussion relies on a proper treatment of measurement disturbance as a genuine causal notion. This approach allows us to make a further distinction between various types thereof, that is, outcome vs parameter dependence represented by the respective arrows $M_1 \to M_2$ and $C \to M_2$ in Fig. 3. For this purpose the full statistics, which includes checking the third box ($C = 1, 2, 3$), appears to be more interesting. It allows us to show that parameter dependence is actually necessary in both pure and realist frameworks [see Figs. 4(c) and 4(d)], i.e.,

$$\text{Full statistics} \underset{C=1,2,3}{\Rightarrow} \text{parameter dependence}.$$

We also showed, by construction of the explicit SCM, that parameter dependence is sufficient to explain the observed statistics (this can be also deduced from the general property regarding saturation of the model by the single arrow $C \to M_2$, as proved in Ref. [33]).

Let us emphasize that a statement that an arrow is unnecessary does not mean that in reality it is not present (this only means that one can explain the statistics without its assistance). However, the necessity of an arrow implies that it cannot be replaced by any other arrow and still correctly reproduce the statistics. In this sense the causal DAGs in Fig. 4 are the minimal structures compatible with the observed statistics under the given realist or pure assumption.

V. DISCUSSION

In this paper we focused on the assessment of the role of the various causal mechanisms capable of generating the statistics observed in the quantum three-box paradox. Such an approach seems to be more revealing regarding the structure of measurement disturbance than the mere acknowledgment of its presence of some sort. The analysis based on the instrumental inequalities shows the necessity of the parameter dependence in the system (but interestingly, only when the full statistics is taken into account). Furthermore, the use of the d-separation criterion allows us to see the paradox as a case of the selection bias [16] (see Refs. [34,35] for a related issue of postselection in the definition of weak values [36]). Note that in our discussion we take the conservative point of view with single measurement outcomes (in contrast to the many-worlds interpretation) as well as exclude backward-in-time causation.

We remark that for the three-box paradox there is no physical principle that would prohibit the propagation of measurement disturbance in the experiment (like the locality principle for Bell experiments [1–3]). This is because the PPS paradigm consists of a sequence of two measurements $\mathcal{M}_1^c$ and $\mathcal{M}_2$, with the second one being made on the whole system (i.e., even if the boxes can be kept separate while checking a given box C in the measurement $\mathcal{M}_1^c$, all of them have to merge together at the end to implement the measurement $\mathcal{M}_2$). See Fig. 1. It is also clearly seen in the experimental realizations of the three-box paradox [37,38]. This makes all the information about the outcome $\mathcal{M}_1^c$ and the parameter C in principle available upon postselection, and hence neither of the disturbance arrows can be excluded by locality arguments. It is in fact possible to construct a generic local hidden variable model of a single particle in arbitrary linear optical circuits (where measurement disturbance propagates locally too) [39].

It is instructive to observe that the quantum description of the paradox, if interpreted literally in the causal terms, works according to the pure causal DAG in Fig. 3 with *both* arrows $M_1 \to M_2$ and $C \to M_2$ present. This is readily seen from the corresponding SCM which boils down to the identification $\Lambda \equiv \rho_0$ and the structural equations implementing the standard quantum recipe. See Appendix C for details. Note that such a quantumlike model is not optimal, as it features the unnecessary arrow $M_1 \to M_2$ (outcome dependence) [see Fig. 4(c) as well as Ref. [33]]. However, it is an interesting open question regarding the scope of scenarios in which outcome dependence can be taken out of the picture in favor of parameter dependence alone (note that in the standard Bell scenario either one of those arrows is just enough).

The aforementioned quantumlike causal model reflects the standard view of quantum theory which turns down the discussion of unobserved properties. It falls within the pure causal setting [see Fig. 3 (left)]. This should be compared with the concept of measurement as an act of observation which reveals some preexisting property of the system. The latter entails including additional structure in the causal modeling of the system as indicated in the realist causal setting [see Fig. 3 (right)]. It is not without consequences for the expressive power of the considered causal structures, which in turn may entail the necessity of certain mechanisms of measurement disturbance in the system (see the comparison in Fig. 4).

Let us note that the PPS paradoxes can be turned into the proofs of contextuality [40–43]. Since in principle all those effects can be attributed to measurement disturbance, it is natural to ask about the possible causal mechanisms allowing for this to be fully attained. This makes the question about the causal resources (here taken as different sorts of arrows) that are enough to explain a given class of contextual effects an interesting research problem. This paper provides a strong hint that parameter dependence is in general indispensable. However, we have also seen that it may be superfluous in some restricted settings (see case $C = 1, 2$ vs $C = 1, 2, 3$ in Fig. 4). We note in passing an interesting question regarding the role of signaling in our argument. Observe that the marginals of M_2 in the behavior in Fig. 2 change only when $C = 3$ is also considered, and it is where the parameter dependence in the three-box statistics can be proved (this should be compared with the nonsignaling Popescu-Rohrlich boxes for which the outcome dependence is just enough). We leave it as a curious remark regarding contextuality in the presence of signaling [44,45].

In conclusion, we mention the existence of several related PPS paradoxes discussed in the literature, e.g., Refs. [46–51]. Their resemblance to the three-box paradox suggests that the causal approach might shed more light there too.

ACKNOWLEDGMENTS

We acknowledge helpful discussions with J. Duda, M. Markiewicz, and R. Staszewski. We especially thank E. Wolfe for bringing Ref. [33] to our attention and comments on the overall framework of this paper.

APPENDIX A: PROOFS OF NECESSITY

Here we answer the question of which of the two red arrows in the causal DAGs in Fig. 3 is necessary in order to reproduce the statistics in Fig. 2. For the proofs we use modern tools of causal inference [16], i.e., the d-separation criterion [Fig. 4(b)] and instrumental inequalities [Figs. 4(c) and 4(d)].

Proof of Fig. 4(b). Let us consider the realist causal framework with just two measurement choices $C = 1, 2$. In that case, from Eq. (4), we have $P(M_1 = 1|M_2 = 1, C) = 1$. As a consequence of the assumption Eq. (5) it means that $V = C$ whenever $M_2 = 1$. This necessitates the conditional dependence $V \not\perp\!\!\!\perp C \,|\, M_2$. However, in the realist causal DAG with both arrows $M_1 \to M_2$ and $C \to M_2$ absent [see Fig. 3 (right)], the variables C and V are d-separated conditioned on M_2 (since the only path $C \to M_1 \leftarrow V$ is blocked by the collider M_1), which entails their statistical independence $V \perp\!\!\!\perp C \,|\, M_2$. This contradiction means that at least one of those arrows must be present in the realist causal DAG. Indeed, either arrow $M_1 \to M_2$ or $C \to M_2$ lifts the d-separation by opening the respective path $V \leftarrow \Lambda \to M_2 \leftarrow M_1 \leftarrow C$ or $V \leftarrow \Lambda \to M_2 \leftarrow C$ (here M_2 is not a collider because of conditioning). This opens a way for both variables to become correlated (whether it is enough with a single arrow is proved by an explicit SCM in Appendix B). In the field of causal inference the phenomenon of correlation due to conditioning is known as a *selection bias* or *Berkson's paradox* [16].

Proof of Figs. 4(c) and 4(d). Now we are concerned with the full choice of measurements $C = 1, 2, 3$. For the proof it is enough to consider the pure causal DAG, i.e., Fig. 4(c) [since this framework is more permissive, the necessity for Fig. 4(c) automatically carries over to the realist case Fig. 4(d)].

Let us reformulate our problem in causal inference terms. Suppose we admit the arrow $M_1 \to M_2$ in the pure causal DAG [see Fig. 3 (left)] and ask about the necessity of the arrow $C \to M_2$. (Note that proving the necessity of arrow $C \to M_2$ in this case will entail its necessity when the arrow $M_1 \to M_2$ is missing, since adding the latter only increases the expressive power of the DAG.) This can be recast as the instrumental scenario [16,52], where the instrument C is used for determining causal influence of M_1 on M_2, both of which are affected by some unobserved (or latent) variable Λ. The crucial assumption in this scenario is the absence of any arrow incoming or outgoing from C except for $C \to M_1$.

It appears that this assumption can be tested by the so-called instrumental inequalities [16,52] which adopted to our case take the form

$$\max_i \sum_j \left[\max_k P(M_1=i, M_2=j|C=k)\right] \leqslant 1, \quad \text{(A1)}$$

where $i, j = 0, 1$ and $k = 1, 2, 3$. Violation of those inequalities by the observed statistics testifies to the presence of some other arrow incoming or outgoing from C (whatever the character of the arrow $M_1 \to M_2$). In our case this could be only the arrow $C \to M_2$, and hence a way to check its necessity.

To unfold the condition Eq. (A1) we observe that it is equivalent to the following set of equations:

$$\begin{aligned}
&P(M_1=0, M_2=0|C=k) + P(M_1=0, M_2=1|C=l) \leqslant 1,\\
&P(M_1=1, M_2=0|C=k) + P(M_1=1, M_2=1|C=l) \leqslant 1,\\
&P(M_1=0, M_2=1|C=k) + P(M_1=0, M_2=0|C=l) \leqslant 1,\\
&P(M_1=1, M_2=1|C=k) + P(M_1=1, M_2=0|C=l) \leqslant 1,
\end{aligned} \quad \text{(A2)}$$

where $kl = 12, 13, 23$. Now it is straightforward to check that the statistics in Fig. 2 violates those inequalities for $kl = 13$ and 23. For example, for $kl = 23$ in the first line in Eq. (A2) we get

$$2/3 + 4/9 = 10/9 > 1. \quad \text{(A3)}$$

Therefore, if the full choice of measurements $C = 1, 2, 3$ is considered, then the arrow $C \to M_2$ must be present in the causal DAG in Fig. 3 (either pure or realist), if it is to reproduce the statistics in Fig. 2 (this single arrow is also enough as proved by an explicit SCM in Appendix B). We note that for the limited choice $C = 1, 2$ (i.e., $kl = 12$) the instrumental inequalities Eq. (A2) remain inviolate, which means that in that case the arrow is unnecessary [in full agreement with Figs. 4(a) and 4(b_1)].

APPENDIX B: PROOFS OF SUFFICIENCY

Here we give a repository of explicit SCMs proving the sufficiency of the considered causal DAGs for each case in Fig. 4. In the following we just specify the structural equations for the respective SCMs and observe that the joint probability distributions $P(M_1=i, M_2=j|C=k)$ are obtained by a straightforward application of the product decomposition (Markov property) for the associated causal DAGs.

Proof. Note that the joint probability distribution generated by the causal DAG in Fig. 4(a) has the following decomposition:

$$\begin{aligned}
&P(M_1=i, M_2=j|C=k)\\
&= \sum_\lambda P(M_1=i|C=k, \Lambda=\lambda)P(M_2=j|\lambda)P(\Lambda=\lambda).
\end{aligned} \quad \text{(B1)}$$

Now, in order to recover the statistics in Fig. 2 for $C = 1, 2$ it is enough to consider a Bernoulli distributed hidden variable $\Lambda \sim \text{Ber}(1/3)$, i.e., we have $\Lambda = 0, 1$ and

$$P(\Lambda = 0) = 2/3, \quad P(\Lambda = 1) = 1/3. \quad \text{(B2)}$$

Then, we set the following structural equations compatible with the diagram in Fig. 4(a):

$$M_1(C, \Lambda) := \Lambda, \quad \text{(B3)}$$

$$M_2(\Lambda, N) := \Lambda N, \quad \text{(B4)}$$

where $N \sim \text{Ber}(1/3)$ is an independent noise variable having a Bernoulli distribution. This defines the SCM, which via Eq. (B1) gives the correct statistics in Fig. 2 for $C = 1, 2$ (only).

Proof of Fig. 4(b_1). Let us start by justifying the sufficiency of the causal DAG in Fig. 4(b_1). In this case, the joint probability distribution has the following decomposition:

$$\begin{aligned}
&P(M_1=i, M_2=j|C=k)\\
&= \sum_{\lambda,v} P(M_1=i|C=k, V=v)P(M_2=j|M_1=i, \Lambda=\lambda)\\
&\qquad P(V=v|\Lambda=\lambda)P(\Lambda=\lambda).
\end{aligned} \quad \text{(B5)}$$

In this case we posit a uniformly distributed hidden variable $\Lambda \sim \text{Uni}(1, 3)$, i.e., we have $\Lambda = 1, 2, 3$ with

$$P(\Lambda = i) = 1/3 \qquad \text{for } i = 1, 2, 3. \quad \text{(B6)}$$

The following structural equations define an SCM compatible with the diagram in Fig. 4(b_1):

$$M_1(C, V) := \delta_{C,V}, \qquad \text{[see Eq. (5)]} \quad \text{(B7)}$$

$$V(\Lambda) := \Lambda, \quad \text{(B8)}$$

$$M_2(M_1, \Lambda) := M_1 N, \quad \text{(B9)}$$

where $N \sim \text{Ber}(1/3)$ is a noise variable with a Bernoulli distribution. This is enough to recover, via Eq. (B5), the statistics in Fig. 2 for $C = 1, 2$ (only).

The sufficiency of the DAG in Fig. 4(b_2) follows immediately from the model in Fig. 4(c) which works for all $C = 1, 2, 3$.

Proof of Fig. 4(c). Here the joint probability distribution generated by the causal DAG in Fig. 4(c) has the following decomposition:

$$\begin{aligned}
&P(M_1=i, M_2=j|C=k)\\
&= \sum_\lambda P(M_1=i|C=k, \Lambda=\lambda)\\
&\qquad P(M_2=j|C=k, \Lambda=\lambda)P(\Lambda=\lambda).
\end{aligned} \quad \text{(B10)}$$

We need an SCM which reconstructs the statistics in Fig. 2 for $C = 1, 2, 3$ (all of them). Consider a uniformly distributed hidden variable $\Lambda \sim \text{Uni}(1, 3)$ which takes three values $\Lambda = 1, 2, 3$. Then we define the following set of structural equations in accord with the diagram in Fig. 4(c):

$$M_1(C, \Lambda) := \delta_{C,\Lambda}, \qquad \text{[see Eq. (5)]} \quad \text{(B11)}$$

$$M_2(C, \Lambda) := \begin{cases} \delta_{C,\Lambda} N & \text{for } C = 1, 2,\\ (1-\delta_{C,\Lambda})(1-N) + \delta_{C,\Lambda} N & \text{for } C = 3, \end{cases} \quad \text{(B12)}$$

where $N \sim \text{Ber}(1/3)$ is a noise variable with a Bernoulli distribution. Such a definition of the SCM recovers, via Eq. (B10), the full statistics in Fig. 2 for all $C = 1, 2, 3$.

Proof of Fig. 4(d). This case is a straightforward extension of the SCM in Fig. 4(c). Here the causal DAG in Fig. 4(d) entails decomposition of the joint probability distribution in the form

$$P(M_1=i, M_2=j|C=k) = \sum_{\lambda,v} P(M_1=i|C=k, V=v)P(M_2=j|C=k, \Lambda=\lambda) P(V=v|\Lambda=\lambda)P(\Lambda=\lambda). \quad \text{(B13)}$$

Following the constriction in Fig. 4(c) we take the hidden variable $\Lambda = 1, 2, 3$ with a uniform distribution $\Lambda \sim \text{Uni}(1, 3)$. Then we introduce an additional variable $V = 1, 2, 3$ and equate it with Λ. This trivial extension provides the required SCM which takes the following explicit form [see Eqs. (B11) and (B12)]:

$$M_1(C, V) := \delta_{C,V}, \quad \text{[see Eq. (5)]} \quad \text{(B14)}$$

$$V(\Lambda) := \Lambda, \quad \text{(B15)}$$

$$M_2(C, \Lambda) := \begin{cases} \delta_{C,\Lambda} N & \text{for } C = 1, 2, \\ (1-\delta_{C,\Lambda})(1-N) + \delta_{C,\Lambda} N & \text{for } C = 3, \end{cases} \quad \text{(B16)}$$

where $N \sim \text{Ber}(1/3)$ is again a Bernoulli noise variable. Those structural equations comply with the diagram in Fig. 4(d) reconstructing, via Eq. (B13), the full statistics in Fig. 2 for all $C = 1, 2, 3$.

APPENDIX C: QUANTUMLIKE CAUSAL MODEL

The formalism of quantum theory in the considered sequential scenario indicates the following. The outcomes of the first measurement $\mathcal{M}_1^C$ are distributed as

$$P(M_1 = i|C) = \text{Tr}\left[\Pi_i^C \rho_0\right]. \quad \text{(C1)}$$

Then the initial quantum state ρ_0 gets updated according to the von Neumann–Lüders rule and the conditional distribution of the outcomes of the second measurement $\mathcal{M}_2$ is given by

$$P(M_2 = j|M_1 = i, C) = \frac{\text{Tr}\left[\Pi_j^{\text{post}} \Pi_i^C \rho_0 \Pi_i^C\right]}{\text{Tr}[\Pi_i^C \rho_0]}. \quad \text{(C2)}$$

Note that this is how Eq. (2) is derived.

Let us construct an SCM as suggested by Eqs. (C1) and (C2). It will have the structure of the pure causal DAG in Fig. 3 (left) with *both* arrows $M_1 \to M_2$ and $C \to M_2$ present, i.e., admit the following decomposition of the joint probability distribution:

$$P(M_1=i, M_2=j|C=k) = \sum_{\lambda} P(M_1=i|C=k, \Lambda=\lambda) P(M_2=j|C=k, M_1=i, \Lambda=\lambda)P(\Lambda=\lambda). \quad \text{(C3)}$$

The obvious choice follows from the trivial identification of the hidden variable Λ with the quantum state $\Lambda \equiv \rho$. In our case, the system is prepared in a given initial state ρ_0, so we have $P(\Lambda = \rho_0) = 1$. Now, the model needs to recover the probabilistic behavior in Eqs. (C1) and (C2). This can be attained with the help of two noise variables N_1 and N_2 with a uniform distribution over interval [0,1]. They are assumed to be independent and correspond, respectively, to M_1 and M_2. Thus we define the following structural equations:

$$M_1(C, \Lambda) := \begin{cases} 0 & \text{if } N_1 < \text{Tr}\left[\Pi_0^C \rho_0\right], \\ 1 & \text{otherwise,} \end{cases} \quad \text{(C4)}$$

and

$$M_2(C, M_1, \Lambda) := \begin{cases} 0 & \text{if } N_2 < \frac{\text{Tr}\left[\Pi_0^{\text{post}} \Pi_{M_1}^C \rho_0 \Pi_{M_1}^C\right]}{\text{Tr}\left[\Pi_{M_1}^C \rho_0\right]}, \\ 1 & \text{otherwise.} \end{cases} \quad \text{(C5)}$$

Clearly, such a defined SCM agrees with the structure of the pure causal DAG in Fig. 3 (left) with *both* arrows $M_1 \to M_2$ and $C \to M_2$ present. Furthermore, the statistics generated by the model, as obtained from Eq. (C3), recovers the desired distribution in Eq. (2) or equivalently Eqs. (C1) and (C2).

[1] J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics* (Cambridge University, Cambridge, England, 1987).

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

[3] V. Scarani, *Bell Nonlocality* (Oxford University, New York, 2019).

[4] S. Kochen and E. Specker, The problem of hidden variables in quantum mechanics, J. Math. Mech. **17**, 59 (1967).

[5] J. Thompson, P. Kurzyński, S.-Y. Lee, A. Soeda, and D. Kaszlikowski, Recent Advances in Contextuality Tests, Open Syst. Inf. Dyn. **23**, 1650009 (2016).

[6] C. Budroni, A. Cabello, O. Gühne, M. Kleinmann, and J. Å. Larsson, Quantum contextuality, arXiv:2102.13036.

[7] Y. Aharonov and L. Vaidman, Complete description of a quantum system at a given time, J. Phys. A **24**, 2315 (1991).

[8] Y. Aharonov, P. G. Bergmann, and J. L. Lebowitz, Time symmetry in the quantum process of measurement, Phys. Rev. **134**, B1410 (1964).

[9] R. E. Kastner, The three-box "paradox" and other reasons to reject the counterfactual usage of the ABL rule, Found. Phys. **29**, 851 (1999).

[10] L. Vaidman, The meaning of elements of reality and quantum counterfactuals: Reply to Kastner, Found. Phys. **29**, 865 (1999).

[11] J. Finkelstein, What is paradoxical about the "Three-box paradox"? arXiv:quant-ph/0606218.

[12] K. A. Kirkpatrick, Classical three-box "paradox," J. Phys. A: Math. Gen. **36**, 4891 (2003).

[13] T. Ravon and L. Vaidman, The three-box paradox revisited, J. Phys. A **40**, 2873 (2007).

[14] K. A. Kirkpatrick, Reply to 'The three-box paradox revisited' by T. Ravon and L. Vaidman, J. Phys. A **40**, 2883 (2007).

[15] O. J. E. Maroney, Measurements, disturbances and the quantum three box paradox, Stud. Hist. Phil. Mod. Phys. **58**, 41 (2017).
[16] J. Pearl, *Causality: Models, Reasoning, and Inference*, 2nd ed. (Cambridge University, Cambridge, England, 2009).
[17] J. Pearl and D. Mackenzie, *The Book of Why: The New Science of Cause and Effect* (Basic, New York, 2018).
[18] P. Spirtes, C. Glymour, and R. Scheines, *Causation, Prediction, and Search*, 2nd ed. (MIT, Cambridge, MA, 2000).
[19] J. Pearl, M. Glymour, and N. P. Jewell, *Causal Inference in Statistics: A Primer* (Wiley, New York, 2016).
[20] M. A. Hernan and J. M. Robins, *Causal Inference: What If* (Chapman and Hall, London, 2020).
[21] J. Peters, D. Janzing, and B. Schölkopf, *Elements of Causal Inference Foundations and Learning Algorithms* (MIT, Cambridge, MA, 2017).
[22] D. B. Rubin and G. W. Imbens, *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction* (Cambridge University, Cambridge, England, 2015).
[23] 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).
[24] R. Chaves, R. Kueng, J. B. Brask, and D. Gross, Unifying Framework for Relaxations of the Causal Assumptions in Bell's Theorem, Phys. Rev. Lett. **114**, 140403 (2015).
[25] K. Ried, M. Agnew, L. Vermeyden, D. Janzing, R. W. Spekkens, and K. J. Resch, A quantum advantage for inferring causal structure, Nat. Phys. **11**, 414 (2015).
[26] M. Ringbauer, C. Giarmatzi, R. Chaves, F. Costa, A. G. White, and A. Fedrizzi, Experimental test of nonlocal causality, Sci. Adv. **2**, e1600162 (2016).
[27] 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).
[28] R. Chaves, G. Carvacho, I. Agresti, V. Di Giulio, L. Aolita, S. Giacomini, and F. Sciarrino, Quantum violation of an instrumental test, Nat. Phys. **14**, 291 (2018).
[29] R. Chaves, G. B. Lemos, and J. Pienaar, Causal Modeling the Delayed-Choice Experiment, Phys. Rev. Lett. **120**, 190401 (2018).
[30] P. Blasiak, E. Borsuk, and M. Markiewicz, On safe post-selection for Bell tests with ideal detectors: Causal diagram approach, Quantum **5**, 575 (2021).
[31] P. Blasiak, E. M. Pothos, J. M. Yearsley, C. Gallus, and E. Borsuk, Violations of locality and free choice are equivalent resources in Bell experiments, Proc. Natl. Acad. Sci. USA **118**, e2020569118 (2021).
[32] J. P. Jarrett, On the physical significance of the locality conditions in the Bell arguments, Nous **18**, 569 (1984).
[33] R. J. Evans, Graphs for margins of Bayesian networks, Scand. J. Statist. **43**, 625 (2016).
[34] C. Ferrie and J. Combes, How the Result of a Single Coin Toss Can Turn Out to be 100 Heads, Phys. Rev. Lett. **113**, 120404 (2014).
[35] L. Vaidman, Comment on "How the result of a single coin toss can turn out to be 100 heads", arXiv:1409.5386.
[36] Y. Aharonov, D. Z. Albert, and L. Vaidman, How the Result of a Measurement of a Component of the Spin of a Spin-1/2 Particle Can Turn Out to be 100, Phys. Rev. Lett. **60**, 1351 (1988).
[37] K. J. Resch, J. S. Lundeen, and A. M. Steinberg, Experimental realization of the quantum box problem, Phys. Lett. A **324**, 125 (2004).
[38] R. E. George, L. M. Robledo, O. J. E. Maroney, M. S. Blok, H. Bernien, M. L. Markham, D. J. Twitchen, J. J. L. Morton, G. A. D. D. Briggs, and R. Hanson, Opening up three quantum boxes causes classically undetectable wavefunction collapse, Proc. Natl. Acad. Sci. USA **110**, 3777 (2013).
[39] P. Blasiak, Local model of a qudit: Single particle in optical circuits, Phys. Rev. A **98**, 012118 (2018).
[40] M. S. Leifer and R. W. Spekkens, Logical pre- and post-selection paradoxes, measurement-disturbance and contextuality, Int. J. Theor. Phys. **44**, 1977 (2005).
[41] M. S. Leifer and R. W. Spekkens, Pre- and Post-Selection Paradoxes and Contextuality in Quantum Mechanics, Phys. Rev. Lett. **95**, 200405 (2005).
[42] M. F. Pusey, Anomalous Weak Values Are Proofs of Contextuality, Phys. Rev. Lett. **113**, 200401 (2014).
[43] M. F. Pusey and M. S. Leifer, Logical pre- and post-selection paradoxes are proofs of contextuality, EPTCS **195**, 295 (2015).
[44] E. N. Dzhafarov, J. V. Kujala, and J.-A. Larsson, Contextuality in three types of quantum-mechanical systems, Found. Phys. **45**, 762 (2015).
[45] J. V. Kujala, E. N. Dzhafarov, and J.-A. Larsson, Necessary and Sufficient Conditions for an Extended Noncontextuality in a Broad Class of Quantum Mechanical Systems, Phys. Rev. Lett. **115**, 150401 (2015).
[46] Y. Aharonov and L. Vaidman, How one shutter can close N slits, Phys. Rev. A **67**, 042107 (2003).
[47] N. Aharon and L. Vaidman, Quantum advantages in classically defined tasks, Phys. Rev. A **77**, 052310 (2008).
[48] L. Vaidman, Past of a quantum particle, Phys. Rev. A **87**, 052104 (2013).
[49] Y. Aharonov, S. Popescu, D. Rohrlich, and P. Skrzypczyk, Quantum Cheshire Cats, New J. Phys. **15**, 113015 (2013).
[50] Y. Aharonov, F. Colombo, S. Popescu, I. Sabadini, D. C. Struppa, and J. Tollaksen, Quantum violation of the pigeonhole principle and the nature of quantum correlations, Proc. Natl. Acad. Sci. USA **113**, 532 (2016).
[51] Y. Aharonov, E. Cohen, A. Landau, and A. C. Elitzur, The case of the disappearing (and re-appearing) particle, Sci. Rep. **7**, 531 (2017).
[52] J. Pearl, On the testability of causal models with latent and instrumental variables, in *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence* (Morgan Kaufmann, 1995), pp. 435–443.

On safe post-selection for Bell tests with ideal detectors: Causal diagram approach

Pawel Blasiak[1], Ewa Borsuk[1], and Marcin Markiewicz[2]

[1]Institute of Nuclear Physics Polish Academy of Sciences, PL-31342 Kraków, Poland
[2]International Centre for Theory of Quantum Technologies, University of Gdańsk, PL-80308 Gdańsk, Poland

Reasoning about Bell nonlocality from the correlations observed in post-selected data is always a matter of concern. This is because conditioning on the outcomes is a source of non-causal correlations, known as a *selection bias*, rising doubts whether the conclusion concerns the actual causal process or maybe it is just an effect of processing the data. Yet, even in the idealised case without detection inefficiencies, post-selection is an integral part of experimental designs, not least because it is a part of the entanglement generation process itself. In this paper we discuss a broad class of scenarios with post-selection on multiple spatially distributed outcomes. A simple criterion is worked out, called the *all-but-one* principle, showing when the conclusions about nonlocality from breaking Bell inequalities with post-selected data remain in force. Generality of this result, attained by adopting the high-level diagrammatic tools of causal inference, provides safe grounds for systematic reasoning based on the standard form of multipartite Bell inequalities in a wide array of entanglement generation schemes, without worrying about the dangers of selection bias. In particular, it can be applied to post-selection defined by single-particle events in each detection chanel when the number of particles in the system is conserved.

1 Introduction

The study of experimental correlations provides a window into the underlying causal mechanisms, even when their exact nature remains obscured. In his seminal works [1], John Bell showed that seemingly innocuous assumptions about the causal structure of realistic models leave a mark on the observed statistics. The conclusion has been that the violation of certain inequalities is incompatible with the assumption of *locality* and *free choice* (or *measurement independence*). Surprisingly, such violations can systematically occur in quantum theory, potentially undermining our dearly held assumptions about how nature works. Given how troubling this conclusion might be, it is hardly surprising how thoroughly Bell's result has been scrutinised in the last few decades, both theoretically [2–4] and experimentally [5]. Its importance is acknowledged by the term *'Bell nonlocality'* which refers to experimental situations demonstrating inconsistency of the observed correlations with causal (or realist) explanations maintaining both assumptions at the same time. It is believed to be a source of quantum advantage in the communication [6] and information tasks [7].

A simplified picture of a Bell experiment consists of a series of measurements made by space-time separated parties on systems prepared in some entangled state. However, it comes with a challenge as to the straight conclusion regarding Bell nonlocality when it comes to the analysis of real experimental designs. An important issue concerns the presence of post-selection in the data collection process. One source of the problem lies with the measurement part of the experiment in which some of the events are missed out due to the inefficiencies of real detectors. It is called the detection loophole [8, 9] and will not be addressed in this paper. In the following we shall assume ideal detectors and thus focus on post-selection

Pawel Blasiak: pawel.blasiak@ifj.edu.pl
Ewa Borsuk: ewa.borsuk@ifj.edu.pl
Marcin Markiewicz: marcinm495@gmail.com

due to the preparation part of the experiment.

It is often the case that post-selection in the experiment is due to the specifics of the entanglement generation process itself. Typically this boils down to the occurrence of a certain pattern in the outcomes deemed to be interesting for the purpose at hand, i.e., exhibiting entanglement [10–12]. Some popular techniques of this sort include: heralding by some other event (cf., event-ready detection [13]), time-bin entanglement [14] or selecting single-particle detections in each experimental channel (cf., recent proposals in Refs. [15–30] or a variety of integrated photonic implementations [12]). So, the generic structure of events is richer than that required for the intended Bell inequality, and post-selection aims at retaining only those experimental trials, based on some well-defined criterion, which are potentially interesting for the violation of the desired inequality. This poses an issue regarding the legitimacy of the conclusion about Bell nonlocality in such scenarios, since conditioning is often a source of non-causal correlations. In the field of causal inference the problem is known as a *selection bias* or *Berkson's paradox* [31–34]. The difficulty being that, in the presence of post-selection, it might be conditioning that leads to correlations breaking Bell inequalities without necessarily claiming Bell nonlocality. How critical it is for the analysis of Bell experiments may attest the effort to close the detection loophole [9], which exploits post-selection due to detector inefficiencies. Here we will assume ideal detectors and focus only on post-selection due to entanglement generation.

It is interesting to ask about general conditions

> *when post-selection, due to entanglement generation, does not compromise the conclusion of nonlocality from the violation of some given Bell inequality.*

So far, this problem has been discussed only for some particular scenarios for two and three parties, and the analysis typically involved the entire pattern of experimental outcomes present in a given experiment [35–37]. In some cases, those issues can be overcome by certain modifications in the experimental arrangements (like for time-bin entanglement in Refs. [38–41]). However, apart from those particular cases no attempt has been made at a general analysis of Bell nonlocality in the presence of post-selection due to the specifics of entanglement generation process. We note that comprehensiveness of such an analysis would require a discussion of both assumptions, *locality* and *free choice*, underlying the derivation of Bell inequalities.

In this paper we give a general criterion for any multipartite scenario, called the *all-but-one* principle, that can be expressed by the following simple intuition:

> *if post-selection can be resolved with one party excluded, then it is safe for Bell nonlocality arguments.*

Meaning that, in such a case, the reasoning based on the standard Bell inequalities is justified despite the issues of post-selection. Crucially, the generality of the result owes to the high-level diagrammatic tools of causal inference honed by Judea Pearl [31–34]. We give a full poof of this criterion preceded with a brief discussion of the selection bias and Bell nonlocality under post-selection.

2 Selection bias and d-separation rules

Post-selection is a procedure of rejecting some of the data from the analysis of an experiment. Technically, it boils down to estimating experimental probabilities subject to some additional condition which depends on the outcomes. It is crucial to make a warning that post-selection is not a harmless procedure, since it is often a source of additional correlations in the retained data. This is potentially dangerous for the task of identification of causal relationships between the variables from the observed correlations. In the field of causal inference the problem is known as a *selection bias* or *Berkson's paradox*.

Let us illustrate the problem with a simple example due to Elwert & Winship [34]. Consider three features of Hollywood actors who could be beautiful B, talented T, and some of them make it to be celebrities C. We may reasonably expect that beauty B and talent T contribute to an actor being considered a celebrity C (pushed to the extreme, imagine that one of these features is enough to become a celebrity), but in general population beauty B and talent T are completely unrelated to one another. Suppose that this is the whole story and hence Fig. 1 (on the left) illustrates the causal diagram behind the data. Now, if we focus on the subpopulation of those actors who made it to the status of celebrities C,

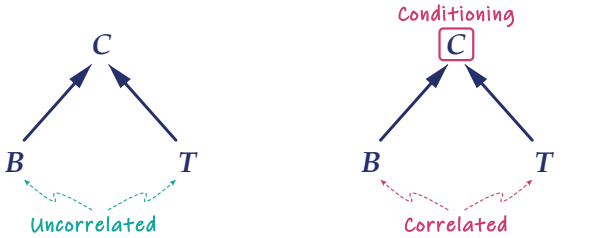


Figure 1: **Selection bias.** Consider three variables B (beauty), T (talent) and C (celebrity) modelled by a causal diagram on the left. It follows that the variables B and T are independent, since the only path joining the variables is blocked by a *collider* C (**Rule 1**). However, on the right, *conditioning* on the collider C, depicted by a red box, unblocks the path between B and T making the variables likely dependent (**Rule 3**).

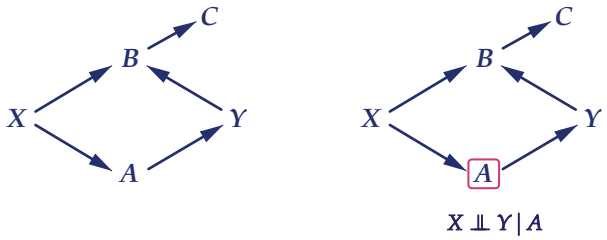


Figure 2: **Illustration of the *d-separation* rules.** In the causal diagram nodes X and Y are connected via two paths $X \to B \leftarrow Y$ and $X \to A \to Y$. The path $X \to B \leftarrow Y$ is blocked since B is a collider by **Rule 1** and conditioning on B or C unblocks the path by **Rule 3**. Whereas path $X \to A \to Y$ is unblocked since A is a non-collider and conditioning on A blocks the path by **Rule 2**. If both paths are blocked then X and Y are said to be *d-separated* and then the variables are independent. In this example it happens only in the case of conditioning on A which is depicted on the right by a red box.

then correlation between beauty B and talent T appears (despite the fact that they were independent to begin with). See Fig. 1 (on the right). Clearly, seeing an unattractive celebrity makes it more likely that the person is a talented actor. And vice versa, celebrities who are bad actors are more often found to be good looking. (Note that pushed to the extreme, this inverse relation may even become a certain conclusion). These correlations are *non-causal*, i.e., they arise merely due to conditioning or restricting the data generated by the causal diagram in Fig. 1 (on the left). This example illustrates the warning against careless attribution of causal origin to correlations in the post-selected data.

Interestingly, the pattern of independencies between the variables can be deduced from the structure of the causal diagram itself. It has been shown to boil down to the so called **d-separation** criterion, see [31–34]. In a nutshell, the idea consists of inspecting all paths in the causal diagram connecting two variables of interest:

> *if all those paths are blocked then the variables are necessarily independent (otherwise the variables are likely dependent).*

The concept of *blocking* a path is defined by the following three simple *d-separation* rules (see Fig. 2 for illustration):

Rule 1. A path is blocked if there is a *collider* along the way, that is a node with pair of arrows on the path that collide head-to-head.

Rule 2. Conditioning on a *non-collider* blocks the path (where non-collider is a node along the way with pair of arrows meeting head-to-tail or tail-to-tail).

Rule 3. Conditioning on a *collider* (or its descendant) removes the block from **Rule 1**.

In the following analysis, those rules will provide systematic insight into the pattern of conditional independences arising from specific post-selection procedures.

3 Bell nonlocality and three causal assumptions

In the following, we consider the usual Bell-type scenario with several parties $A, B, C, ...$ conducting experiments in space-time separated regions. The whole experiment consists of a series of trials in which each party chooses a setting $x, y, z, ...$ and makes a measurement registering the outcome $a, b, c, ...$. For further convenience, let us denote the set of possible outcomes by $\mathcal{O}$, i.e., we have respectively $\mathcal{O}_A$, $\mathcal{O}_B$, $\mathcal{O}_C$, ... Then, after many repetitions, the parties compare their results calculating the statistics given by the set of distributions $P_{abc...|xyz...}$ which describe the probability of observing outcomes $a, b, c, ...$ given measurements $x, y, z, ...$ were made. For conciseness, following the terminology in Ref. [4], we call such obtained set of probabilities $\mathcal{P} \equiv \{P_{abc...|xyz...}\}_{xyz...}$ a *"behaviour"*. Note that all probabilities in the behaviour $\mathcal{P}$ are supposed to be calculated *without* rejecting any trial from the experiment (no post-selection is made).

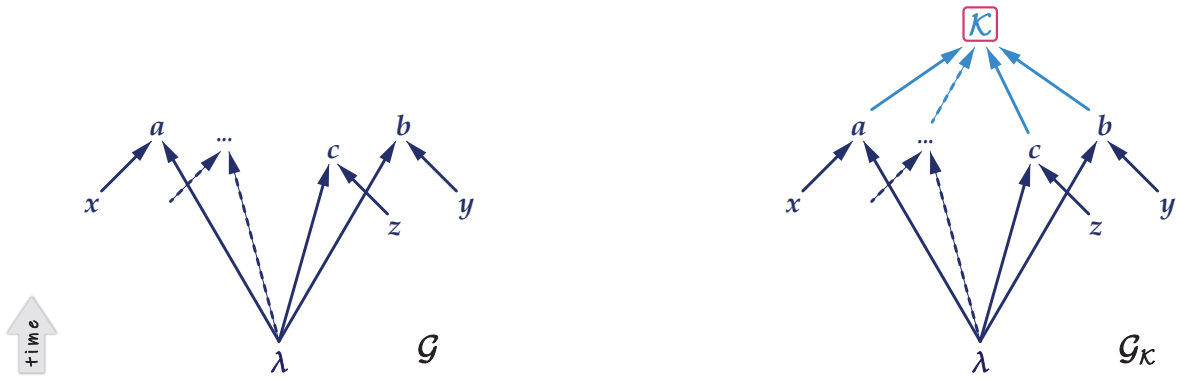


Figure 3: **Causal structure in a Bell experiment.** One the left, the graph $\mathcal{G}$ describes causal relations between variables $a, b, c, ...$ (measurement outcomes), $x, y, z, ...$ (choice of settings), and some hidden variable λ. It generates the behaviour $\mathcal{P} \equiv \{P_{abc...|xyz...}\}_{xyz...}$. On the right, the graph $\mathcal{G}_\mathcal{K}$ incorporates post-selection into the experiment by conditioning (red box) on the additional outcome-dependent variable $\mathcal{K}(a, b, c, ...)$. The latter introduces potential *selection bias* into such obtained behaviour $\mathcal{P}_\mathcal{K} \equiv \{P_{abc...|xyz...\,\mathcal{K}}\}_{xyz...}$.

Bell inequalities are algebraic constraints in the form

$$\mathcal{I}(\mathcal{P}) \equiv \sum_{\substack{abc...\\xyz...}} s^{abc...}_{xyz...}\, P_{abc...|xyz...} \;\leqslant\; I_L\,, \qquad (1)$$

where $s^{abc...}_{xyz...}$ and I_L are some numbers. These inequalities are derived under three assumptions called realism, locality and free choice. The *realism assumption* posits that the observed correlations can be explained by the causal influence between the variables relevant for the experiment, that is measurement outcomes $a, b, c, ...$ and settings $x, y, z, ...$, as well as some unobserved (hidden) variables collectively denoted by λ. Thus, by conditioning on *a priori* unknown λ, we can always write [1–4]

$$P_{abc...|xyz...} = \sum_\lambda\; P_{abc...|xyz...\,\lambda} \cdot P_{\lambda|xyz...}\,. \qquad (2)$$

Then, by invoking spatio-temporal structure of the experiment certain conditional independencies between the variables can be justified. First, the variables in different space locations cannot affect each other and the causal influences propagate respecting temporal order of events. Second, the hidden variable λ is identified to be in the common past of variables representing the outcomes $a, b, c, ...$, but not the variables representing choice of the settings $x, y, z, ...$ and hence the latter cannot be affected by λ. The ensuing causal structure of the variables modelling the experiment is depicted in the causal graph $\mathcal{G}$ in Fig. 3 (on the left). This readily translates into conditional independencies in the statistics generated by those causal models.[1] They are referred to as the *locality* assumption

$$P_{abc...|xyz...\,\lambda} = P_{a|x\lambda} \cdot P_{b|y\lambda} \cdot P_{c|z\lambda} \cdot ...\,, \qquad (3)$$

and the *free choice* assumption (also called the *measurement independence* assumption)

$$P_{\lambda|xyz...} = P_\lambda\,. \qquad (4)$$

Within the causal model framework, these relations are a straightforward application of the *d*-separation rules to the diagram in Fig. 3 (on the left). [Eq. (3) follows by iterative use of **Rule 2** given conditioning on non-collider node λ, and Eq. (4) is an application of **Rule 1** to colliders $a, b, c, ...$; cf. proof of **Theorem 1**.]

To summarise, each Bell inequality Eq. (1) is a simple algebraic consequence of the three assumptions in Eqs. (2)-(4). It means that the violation of some Bell inequality entails the impossibility of explaining the observed behaviour $\mathcal{P}$ in a causal model maintaining both locality and free choice at the same time. The essence of Bell's theorem is to point out situations in which quantum theory predicts violation of those inequalities [1–4].

4 Post-selection issues

Crucially, the statistics used for estimation of probabilities in the behaviour $\mathcal{P}$ should include

[1] Note that in this argument causal relationships are considered as prior to the statistical relations, with the latter derivable from an appropriate structural causal model compatible with a given causal structure. This is the leitmotif of the causal inference field [31–34].

every experimental trial for a valid conclusion from breaking Bell inequality in Eq. (1) to be drawn. In practice, however, some sort of post-selection is always made. Let us formalise this concept by assuming that the causal structure encoded in the causal graph $\mathcal{G}$ remains the same, but admits a richer variety of outcomes $\widetilde{\mathcal{O}}$, i.e., we have $\widetilde{\mathcal{O}}_A \supset \mathcal{O}_A$, $\widetilde{\mathcal{O}}_B \supset \mathcal{O}_B$, $\widetilde{\mathcal{O}}_C \supset \mathcal{O}_C$, ... It means that, in addition to the outcomes of interest $\mathcal{O}$, the experiment predicts results which will have to be rejected in the analysis. Then post-selection boils down to conditioning on some outcome-dependent variable

$$\mathcal{K} \equiv \mathcal{K}(a, b, c, ...)\,. \tag{5}$$

Say, for $\mathcal{K} = 1$ we accept the result, otherwise for $\mathcal{K} = 0$ the result is rejected. This procedure aims at recovering the proper structure of outcomes for the intended Bell inequality Eq. (1). In other words, the reduction $\widetilde{\mathcal{O}} \rightsquigarrow \mathcal{O}$ is achieved by making sure that the unwanted results drop out under the conditioning, i.e., $P_{abc...|xyz...\lambda\mathcal{K}} = 0$ if $a \notin \mathcal{O}_A$ or $b \notin \mathcal{O}_B$ or $c \notin \mathcal{O}_C$, ... Note that the value of $\mathcal{K}$ is decided only after the parties meet to compare their results. Hence the causal graph takes the form $\mathcal{G}_\mathcal{K}$ in Fig. 3 (on the right).

In this way we get a new behaviour $\mathcal{P}_\mathcal{K} \equiv \{P_{abc...|xyz...\mathcal{K}}\}_{xyz...}$ which looks like a good candidate for a test of Bell inequalities Eq. (1). Indeed, all premises seem to be satisfied, i.e., the 'right' causal graph $\mathcal{G}$ with the appropriate pattern of outcomes $\mathcal{O}$, except one detail: there is conditioning in such obtained statistics. This raises worries as regards the validity of the conclusions, reached by using $\mathcal{P}_\mathcal{K}$ in Eq. (1), as a legitimate proof of Bell nonlocality. The selection bias may serve here as a warning of how easily post-selection can lead to false causal conclusions, cf. Fig. 1. In the case of a Bell test it might happen that conditioning (post-selection) bootstraps the correlations so that $\mathcal{I}(\mathcal{P}_\mathcal{K}) > I_L$, while for the full statistics it remains $\mathcal{I}(\mathcal{P}) \leqslant I_L$ in agreement with the causal graph $\mathcal{G}$ in Fig. 3 (i.e., with locality and free choice maintained). This is possible because conditioning ruins the independence structure of Eqs. (3) and (4) which is required to prove Eq. (1). To see this, note that the outcomes $a, b, c, ...$ play the role of colliders in the causal graph $\mathcal{G}$ but conditioning on their descendent $\mathcal{K}$ in graph $\mathcal{G}_\mathcal{K}$ in Fig. 3 opens paths that were previously blocked, cf. **Rule 1** and **Rule 3**, thereby introducing correlations into the data which may fake Bell inequalities. How critical it might be for the analysis of Bell nonlocality may attest the effort to close the detection loophole [3, 4, 8, 9] (which is a case of post-selection too).

Having warned against jumping to hasty conclusions with post-selected data, it is then natural to ask:

> *When is it possible to make a conclusive Bell argument in the post-selected regime?*

In order to make it precise, we assume that the considered causal structure for the experiment is given by the diagram $\mathcal{G}_\mathcal{K}$ in Fig. 3 and make the following definition:

Definition 1 (Safe post-selection)**.**
Post-selection procedure specified by the variable $\mathcal{K}(a, b, c, ...)$ *is considered to be safe if the locality and free choice assumptions still hold in the post-selected regime, i.e.,*

$$P_{abc...|xyz...\lambda\mathcal{K}} = P_{a|x\lambda\mathcal{K}} \cdot P_{b|y\lambda\mathcal{K}} \cdot P_{c|z\lambda\mathcal{K}} \cdot ...\,, \tag{6}$$

and

$$P_{\lambda|xyz...\,\mathcal{K}} = P_{\lambda|\mathcal{K}}\,. \tag{7}$$

An immediate consequence is the observation:

Corollary 1. *If post-selection* $\mathcal{K}$ *is safe, then in the post-selected regime the same set of Bell inequalities holds, i.e.,*

$$\mathcal{I}(\mathcal{P}) \;\leqslant\; I_L \;\Rightarrow\; \mathcal{I}(\mathcal{P}_\mathcal{K}) \;\leqslant\; I_L\,. \tag{8}$$

It follows from the fact that each Bell inequality is obtained by algebraic manipulation of the expression on the l.h.s. of Eq. (1) assuming Eqs. (3) and (4) hold. Clearly, the same must be true for $\mathcal{P}_\mathcal{K}$ since the same algebra, now with Eqs. (6) and (7), must give the same result.

In this way, the problem of validity of reasoning in the post-selected regime with the same Bell inequalities is phrased in terms of conditional independencies in a given causal structure. Given post-selection $\mathcal{K}$, the latter can be efficiently scrutinised with the diagrammatic tools of causal inference (d-separation rules) applied to the causal graph $\mathcal{G}_\mathcal{K}$.

5 Main result

Suppose that the behaviour $\mathcal{P}$ produced by causal graph $\mathcal{G}$ features extra correlations due to the specifics of the preparation procedure (e.g. the

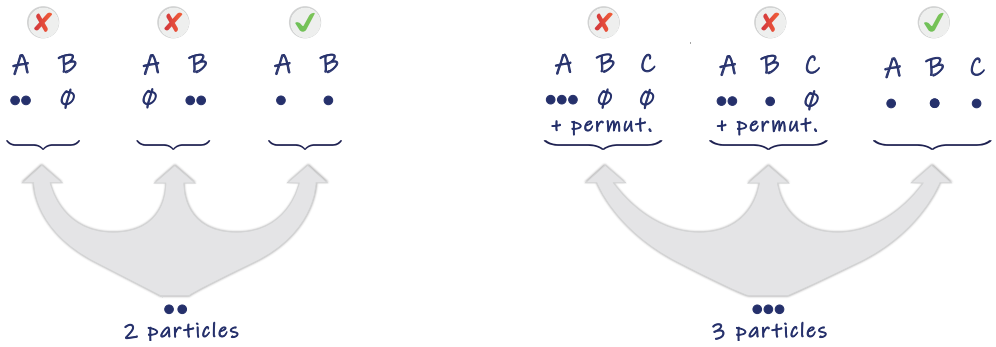


Figure 4: **Illustration of the *all-but-one* principle.** Suppose that in a repeated experiment two particles arrive to Alice (A) and Bob (B) in all possible configurations depicted on the left. Let the meaningful measurements are only those performed on a single particle by each party. If the number of particles is conserved, then Alice alone will know which trials to discard (and similarly for Bob). On the right, there is another party Charlie (C) and three particles are distributed in all possible ways. Again, conservation of particles allows to decide by any two of them AB, AC or BC which trials need to be rejected (without knowing what is happening respectively in C, B or A).

number of particles in the experiment is conserved). This often serves as a means of simplification for post-selection $\mathcal{K}$. Let us consider the following property:

Definition 2 (All-but-one)**.**
Post-selection $\mathcal{K}$ conforms to the all-but-one principle, if it can be fully determined without knowing one of the outcomes. Formally, this boils down to a condition on the form of the variable $\mathcal{K} = \mathcal{K}(a_1, a_2, ..., a_N)$ which requires that it can be reduced to

$$\mathcal{K} \;=\; \mathcal{K}(a_1, \dots, \not{a}_k, \dots, a_N)\,, \tag{9}$$

for each $k = 1, \dots, N$. Here, the symbol $\not{a}_k$ means that the outcome for k-th party is missing.

Informally, this means that in a Bell scenario all parties except one is always enough to know whether post-selection ends with a success or not ($\mathcal{K} = 1$ or 0). For example, for three parties A, B and C, already two of them may decide if post-selection takes place or not, i.e., we have

$$\mathcal{K}(a, b, c) \;=\; \mathcal{K}(b, c) \;=\; \mathcal{K}(a, c) \;=\; \mathcal{K}(a, b)\,. \tag{10}$$

Example. A typical situation where *all-but-one* principle can be readily applied is when the number of particles is conserved. Suppose that N particles are distributed among N parties which receive the particles in different configurations. Let the interesting measurement results are only those when there is a single particle per party. This means that, on top of the valid experimental runs there will be trials in which some parties will register no or more than one particle. Hence the experiment must resort to post-selection which consists of retaining only those trials when each party reports a single particle on their side. Observe that since the total number of particles N is conserved, such a post-selection conforms to the *all-but-one* principle. This is because gathering the outcomes from $N-1$ parties is enough to infer the number of particles received by the missing one (i.e., $N-1$ parties registering a single particle may conclude that the remaining one registers a single particle too, since the total number of particles is N) and hence to resolve post-selection only by themselves. See Fig. 4 for an illustration and Refs. [10–12, 15–26, 26–30] for some experimental designs with post-selection of the *all-but-one* type.

Let us observe that the *all-but-one* principle has non-trivial consequences for the causal graph $\mathcal{G}_\mathcal{K}$ in Fig. 3. Namely, if the statistics generated by graph $\mathcal{G}$ gives a promise of **Definition 2**, then one of the arrows pointing to $\mathcal{K}$ in graph $\mathcal{G}_\mathcal{K}$ can be always erased without in any way affecting the generated statistics (in particular, this means that $\mathcal{P}_\mathcal{K}$ will remain unchanged).

Now we can state our main result:

Theorem 1.
For arbitrary number of parties, post-selection which conforms to the all-but-one principle is always safe.

Proof. Here we sketch the proof for two parties A and B which serves to illustrate the main ideas. For the full proof see Appendix. We need to justify that Eqs. (6) and (7) hold under the following

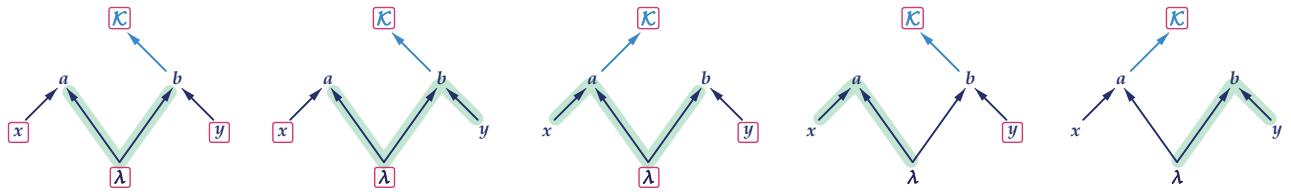


Figure 5: **Graphical proof of Theorem 1 (two parties).** Each graph illustrates structure of conditioning in Eqs. (12) - (13) depicted by red boxes around the variables. Marked in green are paths joining variables of interest for which the respective independencies are inferred by the d-separation **Rules 1 - 3**. In the first three graphs (on the left) conditioning on the non-collider λ blocks the paths, while for the last two graphs (on the right) the paths are blocked by the colliders a and b respectively. Note that without erasing one of the arrows coming to $\mathcal{K}$, as allowed by condition Eq. (11), the inference of conditional independencies would not be possible.

condition:

$$\mathcal{K} = \mathcal{K}(b) = \mathcal{K}(a)\,, \tag{11}$$

i.e., the *all-but-one* principle in **Definition 2**.

As for Eq. (6) the reasoning follows the usual route starting with the standard chain rule, i.e., we have

$$P_{ab|xy\lambda\mathcal{K}} = P_{a|bxy\lambda\mathcal{K}} \cdot P_{b|xy\lambda\mathcal{K}}\,.$$

Then the proof boils down to justification of the following conditional independencies:

$$\begin{aligned} P_{a|bxy\lambda\mathcal{K}} &= P_{a|xy\lambda\mathcal{K}} \quad \text{since} \quad a \perp\!\!\!\perp b \,|\, xy\lambda\mathcal{K}(b) \\ &= P_{a|x\lambda\mathcal{K}} \quad \text{since} \quad a \perp\!\!\!\perp y \,|\, x\lambda\mathcal{K}(b)\,, \\ P_{b|xy\lambda\mathcal{K}} &= P_{b|y\lambda\mathcal{K}} \quad \text{since} \quad b \perp\!\!\!\perp x \,|\, y\lambda\mathcal{K}(a)\,. \end{aligned} \tag{12}$$

Each of them can be inferred from the causal graph $\mathcal{G}_K$ and application of the d-separation **Rule 2** to the non-collider node λ. See Fig. 5 for illustration.

Similarly, we can justify Eq. (7) and get

$$\begin{aligned} P_{\lambda|xy\mathcal{K}} &= P_{\lambda|y\mathcal{K}} \quad \text{since} \quad \lambda \perp\!\!\!\perp x \,|\, y\mathcal{K}(b) \\ &= P_{\lambda|\mathcal{K}} \quad \text{since} \quad \lambda \perp\!\!\!\perp y \,|\, \mathcal{K}(a)\,. \end{aligned} \tag{13}$$

This time it follows from the d-separation **Rule 1**, since a and b are colliders respectively, and **Rule 3** does not apply (note that in neither case $\mathcal{K}$ is a descendent). See Fig. 5 for illustration.

Note that crucial for this line of reasoning is the flexibility of the expression $\mathcal{K}$ in Eq. (11), due to the *all-but-one* principle, which permits to erase one of the arrows coming to $\mathcal{K}$. This trick prevents unblocking certain paths required for the inference of conditional independences. [For example, in last graph on the right in Fig. 5 retaining arrow $b \to \mathcal{K}$ would have opened the (green) path $\lambda \to b \leftarrow x$, since $\mathcal{K}$ becomes then a descendent of the collider b and **Rule 3** applies.]

In general, the complexity of paths that need to be considered in the causal graph $\mathcal{G}_\mathcal{K}$ grows with the number of parties and then the d-separation tools prove indispensable for this kind of analysis, see Appendix. □

6 Discussion

Because entanglement is not a property generated on demand, every Bell experiment must resort to post-selection. However, this opens the doors to the selection bias introducing non-causal correlations into the data, and thus threatening the conclusions expected to be drawn from the experiment. Therefore it is important that the analysis of Bell nonlocality takes this fact into account. In this paper we gave a simple criterion, called the *all-but-one* principle, that allows for safe reasoning in the post-selected regime. Technically, we prove a theorem showing that Bell inequalities derived from the full causal graph which includes conditioning due to post-selection of the *all-but-one* type remain unchanged. It means that the conclusions drawn from breaking Bell inequalities with such a post-selected data remain in full force. Beyond the foundational research and application in multipartite entanglement generation schemes [10–12], this criterion should be significant for quantum cryptography and device independent certification [4, 6].

Novelty of the result reported in this work is three-fold: (a) it concerns *any* multipartite scenario with an arbitrary number of outcomes and settings, (b) it pertains to *any* Bell inequality that can be derived in a given scenario, and

(*c*) *both* assumptions of locality and free choice are explicitly discussed in our analysis. The generality of the *all-but-one* principle should be compared with other treatments of post-selection problem due to entanglement generation [35–37].

Let us emphasise that the *all-but-one* principle draws on a special kind of correlations built into the data due to the specifics of the preparation procedure. Typically, if events of interest consist of arrival of a single particle in each detection channel and the number of particles is known and conserved, then such a post-selection fulfils the *all-but-one* principle. We remark that it is a common situation in quantum optical schemes for entanglement generation, see Refs. [10–12]. Some recent proposals based on coincidence counts for high-dimensional multiparticle entanglement that fall within the *all-but-one* principle include entanglement by path identity [16, 17], entanglement without touching [18–21] or spatial overlap of indistinguishable particles [23, 24]. See also Refs. [25–30]. For completeness, we note that the principle is not applicable to time-bin entanglement scheme [14] which requires specific treatment [38–41].

Note also that the *all-but-one* principle has limitations. Although applicable in many theoretical settings it does not hold in situations with detector inefficiencies (when the number of particles is not predictable). This is a serious matter of concern for experimental tests of Bell inequalities leading to the so called detection loophole which has to be analysed by other means [3, 4, 8, 9]. We also note that an important experimental technique based on event-ready-detection [13] is beyond the scope of the principle in the present form (however, it allows for a straightforward extension to include that scenario too).

We remark that in the paper we take a conservative approach to the analysis of Bell nonlocality with causes propagating forward-in-time. For a discussion of retrocausality see e.g. [42, 43]. Note also that we consider a situation in which both assumptions of locality and free choice are maintained at the same time. For a discussion of partial relaxation of those assumptions see [44] and references therein.

Finally, let us highlight the role of conceptual tools of causal inference [31–34] in the present analysis. Not only this is an inspiring and rigorous framework for the discussion of correlations vs cause-and-effect relations, but comes equipped with the high-level diagrammatic tools (*d*-separation rules) which prove indispensable for the treatment of multipartite Bell scenarios with many observers and outcomes. Despite a fairly recent development of the field of causal inference outside of physics, those methods have already successfully influenced the research in quantum foundations, see e.g. [44–53].

Acknowledgments

We thank M. Żukowski for bringing post-selection issues to our attention. We appreciate Y.-S. Kim and R. Lo Franco for helpful comments. We acknowledge partial support by the Foundation for Polish Science (IRAP project, ICTQT, contract no. MAB/2018/5, co-financed by EU within Smart Growth Operational Programme).

References

[1] J. S. Bell, *Speakable and unspeakable in quantum mechanics* (Cambridge University Press, 1987).

[2] H. M. Wiseman, "The two Bell's theorems of John Bell," J. Phys. A: Math. Theor. **47**, 424001 (2014).

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

[4] V. Scarani, *Bell Nonlocality* (Oxford University Press, 2019).

[5] A. Aspect, "Closing the Door on Einstein and Bohr's Quantum Debate," Physics **8** (2015).

[6] A. Ekert and R. Renner, "The ultimate physical limits of privacy," Nature **507**, 443 (2014).

[7] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2000).

[8] R. D. Gill, "Statistics, Causality and Bell's Theorem," Statist. Sci. **29**, 512 (2014).

[9] J.-A. Larsson, "Loopholes in Bell inequality tests of local realism," J. Phys. A: Math. Gen. **47**, 424003 (2014).

[10] J.-W. Pan, Z.-B. Chen, C.-Y. Lu, H. Weinfurter, A. Zeilinger, and M. Żukowski, "Multiphoton entanglement and interferometry," Rev. Mod. Phys. **84**, 777 (2012).

[11] M. Erhard, M. Krenn, and A. Zeilinger, "Advances in high-dimensional quantum entanglement," Nat. Rev. Phys. **2**, 365 (2020).

[12] J. Wang, F. Sciarrino, A. Laing, and M. G. Thompson, "Integrated photonic quantum technologies," Nat. Photonics **14**, 273 (2020).

[13] M. Żukowski, A. Zeilinger, M. A. Horne, and A. K. Ekert, ""Event-Ready-Detectors" Bell Experiment via Entanglement Swapping," Phys. Rev. Lett. **71**, 4287 (1993).

[14] J. D. Franson, "Bell inequality for position and time," Phys. Rev. Lett. **62**, 2205 (1989).

[15] Q. Zhang, X.-H. Bao, C.-Y. Lu, X.-Q. Zhou, T. Yang, T. Rudolph, and J.-W. Pan, "Demonstration of a scheme for the generation of "event-ready" entangled photon pairs from a single-photon source," Phys. Rev. A **77**, 062316 (2008).

[16] M. Krenn, A. Hochrainer, M. Lahiri, and A. Zeilinger, "Entanglement by Path Identity," Phys. Rev. Lett. **118**, 080401 (2017).

[17] J. Kysela, M. Erhard, A. Hochrainer, M. Krenn, and A. Zeilinger, "Path identity as a source of high-dimensional entanglement," Proc. Natl. Acad. Sci. U.S.A. **117**, 26118 (2020).

[18] P. Blasiak and M. Markiewicz, "Entangling three qubits without ever touching," Sci. Rep. **9**, 20131 (2019).

[19] Y.-S. Kim, T. Pramanik, Y.-W. Cho, M. Yang, S.-W. Han, S.-Y. Lee, M.-S. Kang, and S. Moon, "Informationally symmetrical Bell state preparation and measurement," Opt. Express **26**, 29539 (2018).

[20] Y.-S. Kim, Y.-W. Cho, H.-T. Lim, and S.-W. Han, "Efficient linear optical generation of a multipartite W state via a quantum eraser," Phys. Rev. A **101**, 022337 (2020).

[21] P. Blasiak, E. Borsuk, M. Markiewicz, and Y.-S. Kim, "Efficient linear-optical generation of a multipartite W state," Phys. Rev. A **104**, 023701 (2021).

[22] S. Stanisic, N. Linden, A. Montanaro, and P. S. Turner, "Generating entanglement with linear optics," Phys. Rev. A **96**, 043861 (2017).

[23] B. Bellomo, R. Lo Franco, and G. Compagno, "N identical particles and one particle to entangle them all," Phys. Rev. A **96**, 022319 (2017).

[24] A. Castellini, B. Bellomo, G. Compagno, and R. Lo Franco, "Activating remote entanglement in a quantum network by local counting of identical particles," Phys. Rev. A **99**, 062322 (2019).

[25] H.-S. Zhong, Y. Li, W. Li, L.-C. Peng, Z.-E. Su, Y. Hu, Y.-M. He, X. Ding, W. Zhang, H. Li, L. Zhang, Z. Wang, L. You, X.-L. Wang, X. Jiang, L. Li, Y.-A. Chen, N.-L. Liu, C.-Y. Lu, and J.-W. Pan, "12-Photon entanglement and Scalable Scattershot Boson Sampling with Optimal Entangled-Photon Pairs from Parametric Down-Conversion," Phys. Rev. Lett. **121**, 250505 (2018).

[26] H. Wang, J. Qin, X. Ding, M.-C. Chen, S. Chen, X. You, Y.-M. He, X. Jiang, L. You, Z. Wang, C. Schneider, J. J. Renema, S. Hofling, C.-Y. Lu, and J.-W. Pan, "Boson Sampling with 20 Input Photons and a 60-Mode Interferometer in a 10^{14}-Dimensional Hilbert Space Demonstration," Phys. Rev. Lett. **123**, 250503 (2019).

[27] R. Lo Franco and G. Compagno, "Indistinguishability of Elementary Systems as a Resource for Quantum Information Processing," Phys. Rev. Lett. **120**, 240403 (2018).

[28] F. Nosrati, A. Castellini, G. Compagno, and R. Lo Franco, "Robust entanglement preparation against noise by controlling spatial indistinguishability," npj Quantum Inf. **6**, 39 (2020).

[29] K. Sun, A. Wang, Z.-H. Liu, X.-Y. Xu, J.-S. Xu, C.-F. Li, G.-C. Guo, A. Castellini, F. Nosrati, G. Compagno, and R. Lo Franco, "Experimental quantum entanglement and teleportation by tuning remote spatial indistinguishability of independent photons," Opt. Lett. **23**, 6410 (2020).

[30] M. R. Barros, S. Chin, T. Pramanik, H.-T. Lim, Y.-W. Cho, J. Huh, and Y.-S. Kim, "Entangling Bosons through Particle Indistinguishability and Spatial Overlap," Opt. Express **28**, 38083 (2020).

[31] J. Pearl, *Causality: Models, Reasoning, and Inference*, 2nd ed. (Cambridge University Press, 2009).

[32] P. Spirtes, C. Glymour, and R. Scheines, *Causation, Prediction, and Search*, 2nd ed. (MIT Press, 2000).

[33] J. Pearl, M. Glymour, and N. P. Jewell,

Causal Inference in Statistics: A Primer (Wiley, 2016).

[34] J. Pearl and D. Mackenzie, *The Book of Why: The New Science of Cause and Effect* (Basic Books, 2018).

[35] S. Popescu, L. Hardy, and M. Żukowski, "Revisiting Bell's theorem for a class of down-conversion experiments," Phys. Rev. A **56**, R4353 (1997).

[36] M. Żukowski, "Violations of local realism in multiphoton interference experiments," Phys. Rev. A **61**, 022109 (2000).

[37] F. Sciarrino, G. Vallone, A. Cabello, and P. Mataloni, "Bell experiments with random destination sources," Phys. Rev. A **83**, 032112 (2011).

[38] S. Aerts, P. Kwiat, J.-A. Larsson, and M. Zukowski, "Two-Photon Franson-Type Experiments and Local Realism," Phys. Rev. Lett. **83**, 2872 (1999).

[39] G. Lima, G. Vallone, A. Chiuri, A. Cabello, and P. Mataloni, "Experimental Bell-inequality violation without the postselection loophole," Phys. Rev. A **81**, 040101 (2010).

[40] G. Carvacho, J. Cariñe, G. Saavedra, Á. Cuevas, J. Fuenzalida, F. Toledo, M. Figueroa, A. Cabello, J.-A. Larsson, P. Mataloni, G. Lima, and G. B. Xavier, "Postselection-Loophole-Free Bell Test Over an Installed Optical Fiber Network," Phys. Rev. Lett. **115**, 030503 (2015).

[41] F. Vedovato, C. Agnesi, M. Tomasin, M. Avesani, J.-A. Larsson, G. Vallone, and P. Villoresi, "Postselection-Loophole-Free Bell Violation with Genuine Time-Bin Entanglement," Phys. Rev. Lett. **121**, 190401 (2018).

[42] H. Price, *Time's Arrow and Archimedes' Point: New Directions for the Physics of Time* (Oxford University Press, 1996).

[43] K. B. Wharton and N. Argaman, "Colloquium: Bell's theorem and locally mediated reformulations of quantum mechanics," Rev. Mod. Phys. **92**, 021002 (2020).

[44] P. Blasiak, E. M. Pothos, J. M. Yearsley, C. Gallus, and E. Borsuk, "Violations of locality and free choice are equivalent resources in Bell experiments," Proc. Natl. Acad. Sci. USA **118**, e2020569118 (2021).

[45] 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).

[46] R. Chaves, R. Kueng, J. B. Brask, and D. Gross, "Unifying Framework for Relaxations of the Causal Assumptions in Bell's Theorem," Phys. Rev. Lett. **114**, 140403 (2015).

[47] K. Ried, M. Agnew, L. Vermeyden, D. Janzing, R. W. Spekkens, and K. J. Resch, "A quantum advantage for inferring causal structure," Nature Phys. (2015).

[48] M. Ringbauer, C. Giarmatzi, R. Chaves, F. Costa, A. G. White, and A. Fedrizzi, "Experimental test of nonlocal causality," Sci. Adv. **2**, e1600162 (2016).

[49] J.-M. A. Allen, J. Barrett, D. Horsman, C. M. Lee, and R. W. Spekkens, "Quantum Common Causes and Quantum Causal Models," Phys. Rev. X **7**, 031021 (2017).

[50] R. Chaves, G. B. Lemos, and J. Pienaar, "Causal Modeling the Delayed-Choice Experiment," Phys. Rev. Lett. **120**, 190401 (2018).

[51] 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).

[52] E. G. Cavalcanti, "Classical Causal Models for Bell and Kochen-Specker inequality Violations Require Fine-Tuning," Phys. Rev. X **8**, 021018 (2018).

[53] P. Blasiak and E. Borsuk, "Causal reappraisal of the quantum three box paradox," arXiv: 2107.13937 (2021).

Appendix

Here we prove **Theorem 1** from the main text. For sake of illustration, we start with the case of three parties. Then we give the full proof for any number of parties.

• Proof of **Theorem 1** for *three* parties A, B and C

[The following proof for three parties aims to better illustrate some additional aspects which do not arise in the two-party case. It also serves to emphasise the significance of d-separation tools of causal inference [31–34] for this kind of analysis.]

Proof. Let us consider a Bell experiment with three parties described by the causal graph $\mathcal{G}_{\mathcal{K}}$ in Fig. 6 and assume that post-selection conforms to the *all-but-one* principle in **Definition 2**, i.e., we have

$$\mathcal{K}(a,b,c) \;=\; \mathcal{K}(b,c) \;=\; \mathcal{K}(a,c) \;=\; \mathcal{K}(a,b)\,. \tag{14}$$

Crucially, this property allows erasing one of the three arrows coming to $\mathcal{K}$ without affecting the generated statistics. This trick will be used to infer conditional independencies in the post-selected behaviour $\mathcal{P}_{\mathcal{K}}$.

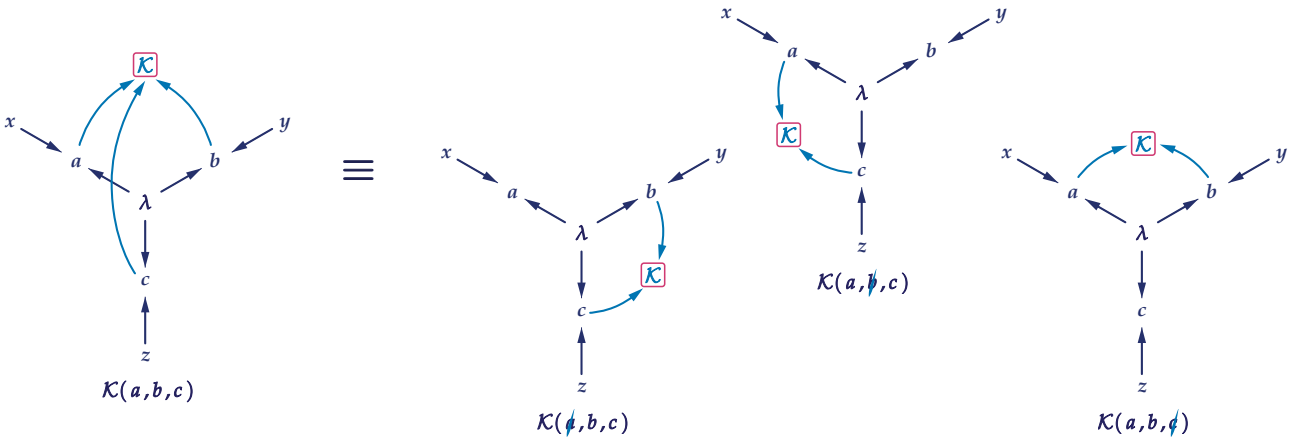


Figure 6: **Causal structure in a Bell experiment for *three* parties with post-selection.** On the left, unfolded on a plain is the causal graph $\mathcal{G}_{\mathcal{K}}$ describing causal relations between variables in a Bell experiment (a, b, c - outcomes, x, y, z - measurement settings, and λ - hidden variable). Variable $\mathcal{K}$ represents post-selection (where the red box means conditioning). Cf. Fig. 3. On the right, three causal graphs which are equivalent to graph $\mathcal{G}_{\mathcal{K}}$, if the *all-but-one* principle in Eq. (14) holds (then one of the three arrows coming to $\mathcal{K}$ can be always erased without affecting the generated statistics).

In order to prove **Theorem 1** we need to justify that both Eqs. (6) and (7) in **Definition 1** hold under the condition in Eq. (14).

As a straightforward application of the chain rule we get

$$P_{abc|xyz\lambda\mathcal{K}} \;=\; P_{a|bcxyz\lambda\mathcal{K}} \cdot P_{bc|xyz\lambda\mathcal{K}} \;=\; P_{a|bcxyz\lambda\mathcal{K}} \cdot P_{b|cxyz\lambda\mathcal{K}} \cdot P_{c|xyz\lambda\mathcal{K}}\,. \tag{15}$$

Now, the prove of Eq. (6) boils down to showing the following sequence of conditional independencies:

$$\begin{aligned} P_{a|bcxyz\lambda\mathcal{K}} &= P_{a|cxyz\lambda\mathcal{K}} && \text{since} \quad a \perp\!\!\!\perp b \,|\, cxyz\lambda\mathcal{K}(b,c) \\ &= P_{a|xyz\lambda\mathcal{K}} && \text{since} \quad a \perp\!\!\!\perp c \,|\, xyz\lambda\mathcal{K}(b,c) \\ &= P_{a|xz\lambda\mathcal{K}} && \text{since} \quad a \perp\!\!\!\perp y \,|\, xz\lambda\mathcal{K}(b,c) \\ &= P_{a|x\lambda\mathcal{K}} && \text{since} \quad a \perp\!\!\!\perp z \,|\, x\lambda\mathcal{K}(b,c)\,. \end{aligned} \tag{16}$$

$$
\begin{aligned}
P_{b|cxyz\lambda\mathcal{K}} &= P_{b|xyz\lambda\mathcal{K}} && \text{since} \quad b \perp\!\!\!\perp c \,|\, xyz\lambda\mathcal{K}(a,c) \\
&= P_{b|yz\lambda\mathcal{K}} && \text{since} \quad b \perp\!\!\!\perp x \,|\, yz\lambda\mathcal{K}(a,c) \\
&= P_{b|y\lambda\mathcal{K}} && \text{since} \quad b \perp\!\!\!\perp z \,|\, y\lambda\mathcal{K}(a,c)\,.
\end{aligned} \tag{17}
$$

$$
\begin{aligned}
P_{c|xyz\lambda\mathcal{K}} &= P_{c|yz\lambda\mathcal{K}} && \text{since} \quad c \perp\!\!\!\perp x \,|\, yz\lambda\mathcal{K}(a,b) \\
&= P_{c|z\lambda\mathcal{K}} && \text{since} \quad c \perp\!\!\!\perp y \,|\, z\lambda\mathcal{K}(a,b)\,.
\end{aligned} \tag{18}
$$

All of them can be inferred by inspecting paths joining variables in question in the causal graph $\mathcal{G}_\mathcal{K}$ and the use of d-separation rules. See Fig. 7. In each case there are two possible paths which are blocked by conditioning on the non-collider node λ (**Rule 2**). Note that in order to get the required conditional independencies all conditions in Eq. (14) need to be used, i.e., the lack of the respective arrow coming to $\mathcal{K}$ in each of the graphs in Fig. 7 is essential.

In a similar manner we can prove Eq. (7), that is we can prove the following conditional independencies:

$$
\begin{aligned}
P_{\lambda|xyz\lambda\mathcal{K}} &= P_{\lambda|yz\mathcal{K}} && \text{since} \quad \lambda \perp\!\!\!\perp x \,|\, yz\mathcal{K}(b,c) \\
&= P_{\lambda|z\mathcal{K}} && \text{since} \quad \lambda \perp\!\!\!\perp y \,|\, z\mathcal{K}(a,c) \\
&= P_{\lambda|\mathcal{K}} && \text{since} \quad \lambda \perp\!\!\!\perp z \,|\, \mathcal{K}(a,b)\,.
\end{aligned} \tag{19}
$$

Now there is always one path joining the relevant variables, and in each case it is blocked by the respective collider a, b and c (**Rule 1**). See Fig. 7. Notice that, here as well, having all three arrows coming to $\mathcal{K}$ would spoil the proof, since it would lift the block from the respective collider by conditioning on its descendent $\mathcal{K}$ (**Rule 3**).

□

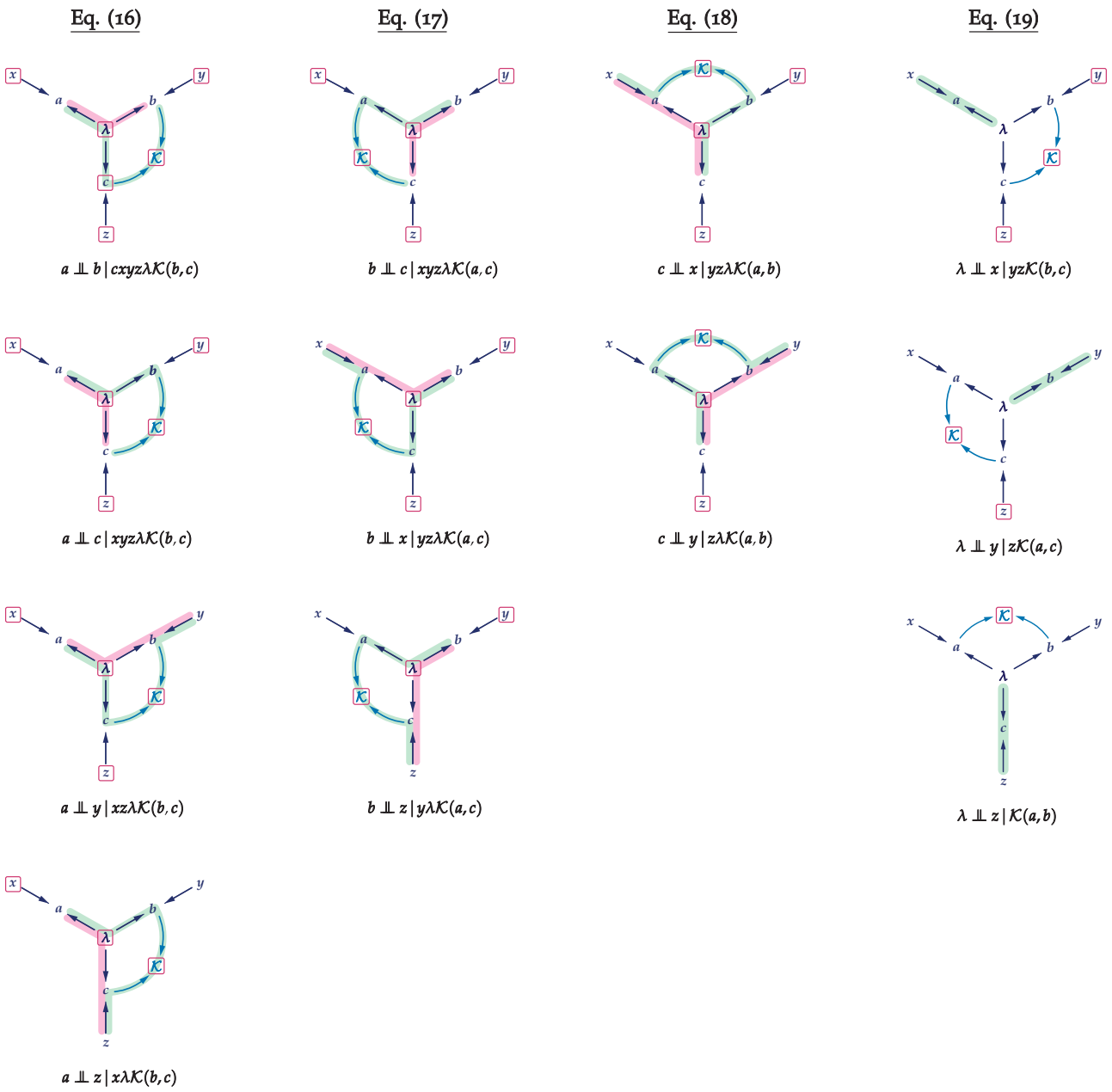


Figure 7: **Graphical proof of Theorem 1 (*three* parties).** Each graph illustrates structure of conditioning in Eqs. (16) - (19) depicted by red boxes around the variables. Marked in purple and green are paths joining variables of interest for which the respective independencies are inferred by the d-separation **Rules 1 - 3**. In case of Eqs. (16) - (18) there are always two paths, and all of them are blocked by conditioning on the non-collider λ (**Rule 2**). As for Eq. (19), in each case the only path is blocked by the respective collider a, b and c (**Rule 1**). Note that dropping one of the arrows in each of the graphs, as allowed by condition in Eq. (14) and explained in Fig. 6, is not accidental (since otherwise it would open additional paths and spoil the independence pattern).

- Proof of **Theorem 1** for *any* number of parties

[The proof follows the lines of reasoning for the case of two and three parties.]

Proof. We consider the general case of N parties in a Bell experiment with post-selection conforming to the *all-but-one* principle in **Definition 2**, i.e we have

$$\mathcal{K} \;=\; \mathcal{K}(a_1,\dots,\not{a}_k,\dots,a_N) \qquad \text{for each} \quad k=1,\dots,N\,, \tag{20}$$

where $\not{a}_k$ means that k-th variable is missing from the list of all outcomes. The causal structure is then given by the graph $\mathcal{G}_{\mathcal{K}}$ in Fig. 8. Note that condition Eq. (20) entails that one of the N arrows coming to $\mathcal{K}$ can be always erased without changing the generated statistics (cf. Fig. 6). In particular, this means that conditional independencies inferred by dropping one of those arrows remain the same.

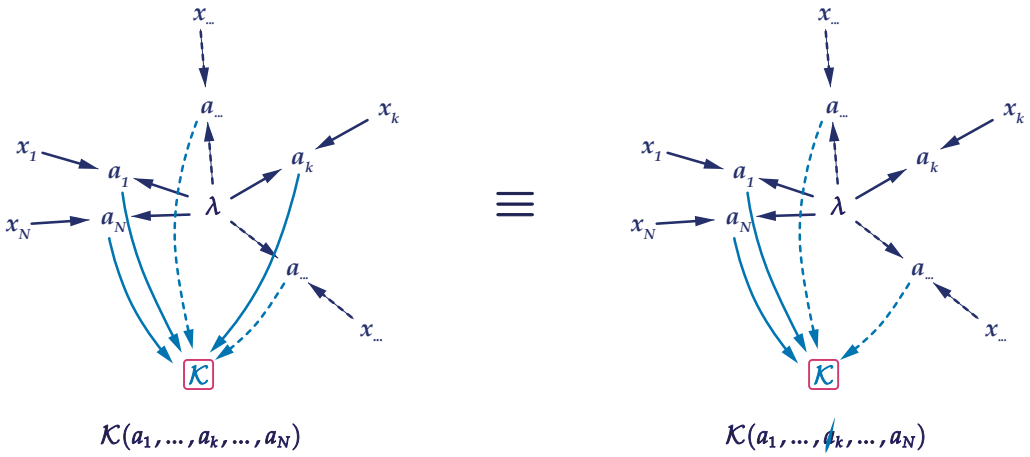


Figure 8: **Causal structure in a Bell experiment for *any* number parties with post-selection.** On the left, unfolded on a plain is the causal graph $\mathcal{G}_{\mathcal{K}}$ describing relations between variables in a Bell experiment ($a_1,\dots,a_N$ - outcomes, $x_1,\dots,x_N$ - measurement settings, and λ - hidden variable). Variable $\mathcal{K}$ represents post-selection (where the red box means conditioning). Cf. Fig. 3. On the right, the causal graph with one arrow $a_k \to \mathcal{K}$ erased which is equivalent to graph $\mathcal{G}_{\mathcal{K}}$, if the *all-but-one* principle in Eq. (20) holds.

The proof of **Theorem 1** consists of justifying that both Eqs. (6) and (7) in **Definition 1** hold under the condition in Eq. (20).

By the repeated use of the chain rule, we get

$$P_{a_1\dots a_N|x_1\dots x_N\lambda\mathcal{K}} \;=\; \prod_{k=1}^{N} P_{a_k|a_{k+1}\dots a_N x_1\dots x_N\lambda\mathcal{K}}\,. \tag{21}$$

In order to prove Eq. (6) it remains to check that for each $k=1,\dots,N$ the following sequence of conditional independencies holds:

$$\begin{aligned}
P_{a_k|a_{k+1}\dots a_N\, x_1\dots x_N\lambda\mathcal{K}} \;&=\; P_{a_k|a_{k+2}\dots a_N\, x_1\dots x_N\lambda\mathcal{K}} \\
&\qquad \text{since} \quad a_k \perp\!\!\!\perp a_{k+1}\,|\,a_{k+2}\dots a_N\, x_1\dots x_N\lambda\,\mathcal{K}(a_1,\dots,\not{a}_k,\dots,a_N) \\
&=\; \dots \\
&=\; P_{a_k|a_{l+1}\dots a_N\, x_1\dots x_N\lambda\mathcal{K}} \\
&\qquad \text{since} \quad a_k \perp\!\!\!\perp a_{l}\,|\,a_{l+1}\dots a_N\, x_1\dots x_N\lambda\,\mathcal{K}(a_1,\dots,\not{a}_k,\dots,a_N) \\
&=\; \dots \\
&=\; P_{a_k|x_1\dots x_N\lambda\mathcal{K}} \\
&\qquad \text{since} \quad a_k \perp\!\!\!\perp a_{N}\,|\,x_1\dots x_N\lambda\,\mathcal{K}(a_1,\dots,\not{a}_k,\dots,a_N)\,,
\end{aligned} \tag{22}$$

where $l > k$, and

$$\begin{aligned}
P_{a_k|x_1\dots x_N\lambda\mathcal{K}} &= P_{a_k|x_2\dots x_N\lambda\mathcal{K}} && \text{since} \quad a_k \perp\!\!\!\perp x_1 \,|\, x_2\dots x_N\lambda\mathcal{K}(a_1,\dots,\not{a}_k,\dots,a_N)\\
&= \dots\\
&= P_{a_k|x_{m+1}\dots x_N\lambda\mathcal{K}} && \text{since} \quad a_k \perp\!\!\!\perp x_m \,|\, x_{m+1}\dots x_N\lambda\mathcal{K}(a_1,\dots,\not{a}_k,\dots,a_N)\\
&= \dots && (23)\\
&= P_{a_k|x_k\, x_{l+1}\dots x_N\lambda\mathcal{K}} && \text{since} \quad a_k \perp\!\!\!\perp x_l \,|\, x_k\, x_{l+1}\dots x_N\lambda\mathcal{K}(a_1,\dots,\not{a}_k,\dots,a_N)\\
&= \dots\\
&= P_{a_k|x_k\lambda\mathcal{K}} && \text{since} \quad a_k \perp\!\!\!\perp x_N \,|\, x_k\lambda\mathcal{K}(a_1,\dots,\not{a}_k,\dots,a_N)\,,
\end{aligned}$$

where $l > k > m$. These independencies can be justified by the d-separation tools applied to the causal graph $\mathcal{G}_\mathcal{K}$. It boils down to the inspection of all paths joining variables in question. See Fig. 9 (cf. Fig. 7). In each case there are $N-1$ possible paths which are all blocked by conditioning on the non-collider node λ (**Rule 2**). Like before, we use all conditions in Eq. (20) to get the results (as readily seen from Fig. 9, where in each case a different arrow $a_k \to \mathcal{K}$ is missing).

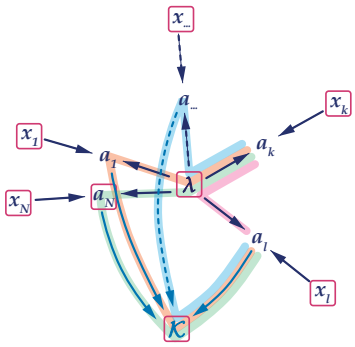
Similarly, the proof of Eq. (7) boils down to the following conditional independencies:

$$\begin{aligned}
P_{\lambda|x_1\dots x_N\mathcal{K}} &= P_{\lambda|x_2\dots x_N\mathcal{K}} && \text{since} \quad \lambda \perp\!\!\!\perp x_1 \,|\, x_2\dots x_N\mathcal{K}(\not{a}_1,a_2,\dots,a_N)\\
&= \dots\\
&= P_{\lambda|x_{k+1}\dots x_N\mathcal{K}} && \text{since} \quad \lambda \perp\!\!\!\perp x_k \,|\, x_{k+1}\dots x_N\mathcal{K}(a_1,\dots,\not{a}_k,\dots,a_N) && (24)\\
&= \dots\\
&= P_{\lambda|\mathcal{K}} && \text{since} \quad \lambda \perp\!\!\!\perp x_N \,|\, \mathcal{K}(a_1,\dots,a_{N-1},\not{a}_N)\,.
\end{aligned}$$

Here there is only one path joining the relevant variables that is blocked by the respective collider a_k (**Rule 1**). See Fig. 9 (cf. Fig. 7). Note that the lacking arrow $a_k \to \mathcal{K}$ is crucial, since otherwise it would unblock the respective collider a_k by conditioning on its descendent $\mathcal{K}$ (**Rule 3**).

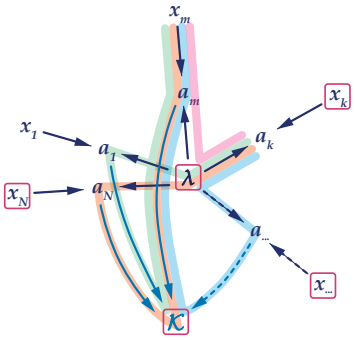
□

Eq. (22)



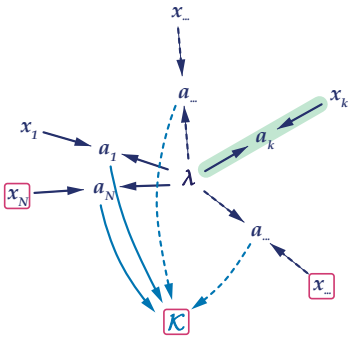
$a_k \perp\!\!\!\perp a_l \,|\, a_{l+1} \ldots a_N\, x_1 \ldots x_N \lambda \mathcal{K}(a_1, \ldots, \not{a_k}, \ldots, a_N)$

Eq. (23)

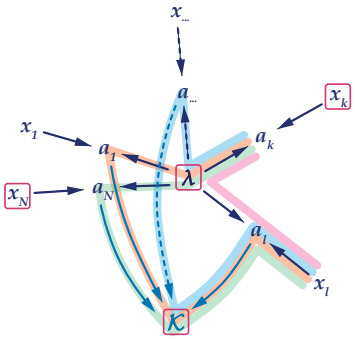


$a_k \perp\!\!\!\perp x_m \,|\, x_{m+1} \ldots x_N \lambda\, \mathcal{K}(a_1, \ldots, \not{a_k}, \ldots, a_N)$

Eq. (24)



$\lambda \perp\!\!\!\perp x_k \,|\, x_{k+1} \ldots x_N \mathcal{K}(a_1, \ldots, \not{a_k}, \ldots, a_N)$



$a_k \perp\!\!\!\perp x_l \,|\, x_k\, x_{l+1} \ldots x_N \lambda\, \mathcal{K}(a_1, \ldots, \not{a_k}, \ldots, a_N)$

Figure 9: **Graphical proof of Theorem 1 (*any* number of parties).** We give three generic cases used for the justification of conditional independencies in Eqs. (22) - (24). Red boxes around the variables depict the structure of conditioning. Marked in purple, green, orange and blue are paths joining variables of interest, a_k and a_l (resp. x_l), for which the respective independencies are inferred by the d-separation **Rules 1 - 3**. In case of Eqs. (22) and (23) there are $N-1$ paths, and all of them are blocked by conditioning on the non-collider λ (**Rule 2**). As for Eq. (24), the only path joining λ and x_k is blocked by the collider a_k (**Rule 1**). Let us note that the lack of the respective arrow $a_k \to \mathcal{K}$ in each of the graphs is essential (in order not to introduce unwanted paths in case of Eqs. (22) and (23), or to prevent lifting the block from the collider by **Rule 3** in case of Eq. (24)).

Violations of locality and free choice are equivalent resources in Bell experiments

Pawel Blasiak[a,b,1], Emmanuel M. Pothos[b], James M. Yearsley[b], Christoph Gallus[c], and Ewa Borsuk[a]

[a]Division of Theoretical Physics, Institute of Nuclear Physics Polish Academy of Sciences, PL-31342 Krakow, Poland; [b]Psychology Department, City, University of London, London EC1V 0HB, United Kingdom; and [c]THM Business School, Technische Hochschule Mittelhessen, D-35390 Giessen, Germany

Edited by Anthony Leggett, University of Illinois at Urbana–Champaign, Urbana, IL, and approved March 11, 2021 (received for review October 1, 2020)

Bell inequalities rest on three fundamental assumptions: realism, locality, and free choice, which lead to nontrivial constraints on correlations in very simple experiments. If we retain realism, then violation of the inequalities implies that at least one of the remaining two assumptions must fail, which can have profound consequences for the causal explanation of the experiment. We investigate the extent to which a given assumption needs to be relaxed for the other to hold at all costs, based on the observation that a violation need not occur on every experimental trial, even when describing correlations violating Bell inequalities. How often this needs to be the case determines the degree of, respectively, locality or free choice in the observed experimental behavior. Despite their disparate character, we show that both assumptions are equally costly. Namely, the resources required to explain the experimental statistics (measured by the frequency of causal interventions of either sort) are exactly the same. Furthermore, we compute such defined measures of locality and free choice for any nonsignaling statistics in a Bell experiment with binary settings, showing that it is directly related to the amount of violation of the so-called Clauser–Horne–Shimony–Holt inequalities. This result is theory independent as it refers directly to the experimental statistics. Additionally, we show how the local fraction results for quantum-mechanical frameworks with infinite number of settings translate into analogous statements for the measure of free choice we introduce. Thus, concerning statistics, causal explanations resorting to either locality or free choice violations are fully interchangeable.

locality | free choice | causality | Bell inequalities | measure of locality and free choice

> I would rather discover one true cause than gain the kingdom of Persia.
>
> Democritus (*ca*. 460–370 BC)

The study of experimental correlations provides a window into the underlying causal mechanisms, even when their exact nature remains obscured. In his seminal works (1–5) John Bell showed that seemingly innocuous assumptions about the structure of causal relationships leave a mark on the observed statistics. The first assumption, called realism (or counterfactual definiteness), presents the worldview in which physical objects and their properties exist, whether they are observed or not. Note that realism allows a standard notion of causality (6, 7), which in turn provides us with the language to express the remaining two assumptions. The locality assumption is a statement that physical (or causal) influences propagate in accord with the spatiotemporal structure of events (i.e., neither backward in time nor instantaneous causation). The free-choice assumption asserts that the choice of measurement settings can be made independently from anything in the (causal) past. These three assumptions are enough to derive testable constraints on correlations called Bell inequalities.

Surprisingly, nature violates Bell inequalities (8–15), which means that if the standard causal (or realist) picture is to be maintained at least one of the remaining two assumptions, that is locality or free choice, has to fail. It turns out that rejecting just one of those two assumptions is always enough to explain the observed correlations, while maintaining consistency with the causal structure imposed by the other. Either option poses a challenge to deep-rooted intuitions about reality, with a full range of viable positions open to serious philosophical dispute (16–18). Notably, quantum theory in its operational formulation does not provide any clue regarding the causal structure at work, leaving such questions to the domain of interpretation. It is therefore interesting to ask about the extent to which a given assumption needs to be relaxed, if we insist on upholding the other one (while always maintaining realism). In this paper, we seek to compare the cost of locality and free choice on an equal footing, without any preconceived conceptual biases. As a basis for comparison we choose to measure the weight of a given assumption in terms of the following question: *How often can a given assumption, i.e., locality or free choice, be retained, while safeguarding the other assumption, in order to fully reproduce some given experimental statistics within a standard causal (or realist) approach?*

This question presumes that a Bell experiment is performed trial-by-trial and the observed statistics can be explained in the standard causal model (or hidden variable) framework (1–7, 19–21), which subsumes realism. It means that the remaining two assumptions of locality and free choice translate into conditional independence between certain variables in the model,

Significance

Faced with a violation of Bell inequalities, a committed realist might pursue an explanation of the observed correlations on the basis of violations of the locality or free choice (sometimes called measurement independence) assumptions. The question of whether it is better to abandon (partially or completely) locality or free choice has been strongly debated since the inception of Bell inequalities, with ardent supporters on either side. We offer a comprehensive treatment that allows a comparison of both assumptions on an equal footing, demonstrating a deep interchangeability. This both advances the foundational debate and provides quantitative answers regarding the weight of each assumption for causal (or realist) explanations of observed correlations.

Author contributions: P.B. designed and performed the research; P.B., E.M.P., J.M.Y., C.G., and E.B. discussed the results and wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: pawel.blasiak@ifj.edu.pl.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2020569118/-/DCSupplemental.

Published April 22, 2021.

PHYSICS

whose causal structure is determined by their spatiotemporal relations (6, 22) (for some alternative approach endorsing indefinite causal structures see, e.g., refs. 23 and 24 or ref. 25 for discussion of retrocausality). Modeling of the experiment implies that in each run of the experiment all variables (including unobserved or hidden ones) always take definite values and the statistics accumulate over many trials. This leaves open the possibility that the violation of the assumptions does not have to occur on each run of the experiment to explain the given statistics. We can thus put flesh on the bones of the above question and seek the maximal proportion of trials in which a given assumption can be retained, while safeguarding the other assumption, so as to fully reproduce some given statistics. In the following, we shall denote so defined measure of locality (safeguarding freedom of choice) as μ_L and measure of free choice (safeguarding locality) as μ_F. Also, without stating this in every instance, we note that in all subsequent discussion realism is assumed.*

There has been some previous research on this theme. A measure of locality analogous to μ_L was first proposed by Elitzur et al. (27) to quantify nonlocality in a singlet state. Note that it seems that the original idea of a bound for such a locality measure was expressed earlier, in ref. 28, but a bound was not worked out. In any case, Elitzur et al.'s measure was dubbed "local fraction" (or "content") and shown (with improvements in refs. 29–31) to vanish in the limit of an infinite number of measurement settings. A substantial step was made in ref. 32, where the local fraction is explicitly calculated for any pure two-qubit state for an arbitrary choice of settings. We note that those results concern measure μ_L only for the specific case of quantum-mechanical predictions. In this paper we go beyond this framework and consider the case of general experimental statistics (see ref. 33 for extension to the idea of contextuality). To avoid confusion, the term "local fraction" for measure μ_L will be only used in relation to the quantum case. Furthermore, we propose a similar treatment of the free-choice assumption quantified by measure μ_F. Natural as it may seem, this approach has not been pursued in the literature, with some other measures proposed to this effect (34–42) [all retaining locality as a principle, but departing from the original notion of free choice introduced by Bell (6, 22)].

We aim to comprehensively consider the extent to which a given assumption, i.e., locality or free choice, can be preserved through partial violation of the other assumption. To accomplish this, we provide similar definitions and discuss on an equal footing both measures of locality μ_L and free choice μ_F. Then, we derive the following results. First, we prove a general structural theorem about causal models explaining any given experimental statistics in a Bell experiment (for any number of settings) showing that such defined measures are necessarily equal, $\mu_L = \mu_F$. This result consolidates those two disparate concepts, demonstrating their deep interchangeability. Second, we explicitly compute both measures for any nonsignaling statistics in a two-setting and two-outcome Bell scenario. This enables a direct interpretation to the amount of violation of the Clauser–Horne–Shimony–Holt (CHSH) inequalities (43). Third, we consider the special case of the quantum statistics with infinite number of settings, utilizing existing results for the local fraction μ_L, which thus translate on the newly developed concept of the measure of free choice μ_F. Fig. 1 summarizes the results in the paper.

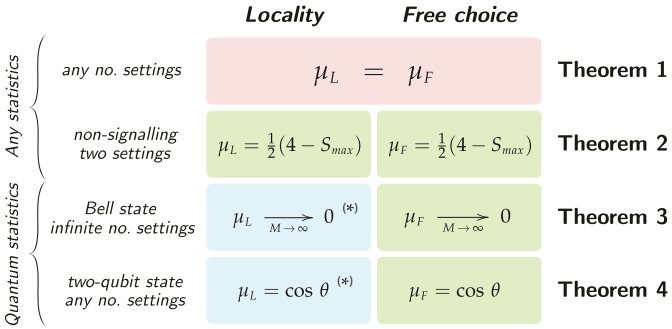


Fig. 1. Summary of the results. The main *Theorem 1* is the backbone of the paper, consolidating both measures of locality μ_L and free choice μ_F. *Theorem 2* is a theory-independent result about both measures μ_L and μ_F. It offers a concrete interpretation for the amount of violation of the CHSH inequalities. *Theorems 3* and *4* are specific to the quantum-mechanical statistics stated here for measure μ_F. They are translations of some remarkable local fraction results μ_L in the literature (marked with an (*); cf. refs. 29–32).

Results

Bell Experiment and Fine's Theorem. Let us consider the usual Bell-type scenario with two parties, called Alice and Bob, playing the roles of agents conducting experiments on two separated systems (whose nature is irrelevant for the argument). We assume that on each side there are two possible outcomes labeled respectively $a, b = \pm 1$ and M possible measurement settings labeled respectively $x, y \in \mathfrak{M}$, where $\mathfrak{M} \equiv \{1, 2, \ldots, M\}$. A Bell experiment consists of a series of trials in which Alice and Bob each choose a setting and make a measurement registering the outcome. After many repetitions, they compare their results described by the set of $M \times M$ distributions $\{P_{ab|xy}\}_{xy}$, where $P_{ab|xy}$ denotes the probability of obtaining outcomes a, b, given measurements x, y were made on Alice and Bob's side, respectively. For conciseness, following the terminology in ref. 5, we will call $\{P_{ab|xy}\}_{xy}$ a "behavior." Note that without assuming anything about the causal structure underlying the experiment any behavior is admissible (as long as the distributions are normalized, i.e., $\sum_{a,b} P_{ab|xy} = 1$ for each $x, y \in \mathfrak{M}$). In particular, quantum theory gives a prescription for calculating the experimental statistics $P_{ab|xy}$ for each choice of settings $x, y \in \mathfrak{M}$ based on the formalism of Hilbert spaces.

It is instructive to recall the special case of two measurement settings on each side $x, y \in \mathfrak{M} = \{0, 1\}$ for which Bell derived his seminal result. Briefly, this can be expressed by saying that any local hidden-variable model with free choice has to satisfy the following four CHSH inequalities (43):

$$|S_i| \leqslant 2 \qquad \text{for } i = 1, \ldots, 4, \qquad \textbf{[1]}$$

where

$$S_1 = \langle ab\rangle_{00} + \langle ab\rangle_{01} + \langle ab\rangle_{10} - \langle ab\rangle_{11}, \qquad \textbf{[2]}$$
$$S_2 = \langle ab\rangle_{00} + \langle ab\rangle_{01} - \langle ab\rangle_{10} + \langle ab\rangle_{11}, \qquad \textbf{[3]}$$
$$S_3 = \langle ab\rangle_{00} - \langle ab\rangle_{01} + \langle ab\rangle_{10} + \langle ab\rangle_{11}, \qquad \textbf{[4]}$$
$$S_4 = -\langle ab\rangle_{00} + \langle ab\rangle_{01} + \langle ab\rangle_{10} + \langle ab\rangle_{11}, \qquad \textbf{[5]}$$

with $\langle ab\rangle_{xy} = \sum_{a,b} ab\, P_{ab|xy}$ being correlation coefficients for a given choice of settings x, y. Interestingly, by virtue of Fine's theorem (44, 45), this is also a sufficient condition for a nonsignaling behavior $\{P_{ab|xy}\}_{xy}$ to be explained by a local hidden-variable model with freedom of choice (for nonsignaling see Eqs. **16** and **17**).

It is crucial to observe that, although locality and freedom of choice are two disparate concepts with different ramifications

*As noted, realism is subsumed in the standard notion of causality, which is implicit in the definition of locality and free choice (1–6). So, henceforth, referring to the standard causal framework implies the realist approach. We also remark that, although, "realism" goes under different guises in the literature (e.g., "counterfactual definiteness," "local causality," "hidden causes," etc.), for our purposes those distinctions are irrelevant and the underlying mathematics remains the same, i.e., it boils down to the hidden-variable framework [which beyond physics is frequently referred to as the structural causal models (7)]. See refs. 3 and 26 for some discussion.

for our understanding of the experiment, they are in a certain sense interchangeable. If locality is dropped with Alice and Bob freely choosing their settings, then the boxes, by influencing one another, can produce any behavior $\{P_{ab|xy}\}_{xy}$. Similarly, a violation of the free-choice assumption can be used to reproduce any behavior $\{P_{ab|xy}\}_{xy}$, without giving up locality. It is straightforward to see how this might work if one of the two assumptions fails on every experimental trial.[†]

However, such a complete renouncement of assumptions so central to our view of nature may seem excessive, especially when the CHSH inequalities are violated only by a little amount (less than the maximal algebraic bound of $|S_i| \leqslant 4$), leaving room for a possible explanation of the experimental statistics by rejecting one of the assumptions sometimes only. Here we assess the cost of such a partial violation by asking how often a given assumption can be retained in order to account for a behavior $\{P_{ab|xy}\}_{xy}$. We will investigate both cases in parallel: [♠] full freedom of choice with occasional nonlocality (communication) and [♣] the possibility of retaining full locality at a price of compromising freedom of choice (by controlling or rigging measurement settings) on some of the trials. We shall use the least frequency of violation, required to model some statistics with a hypothetical simulation, as a natural figure of merit, guided by the principle that the less the violation the better. Notably, such simulations should not restrict possible distributions of measurement settings P_{xy}. In other words, we define a measure of locality μ_L as

> *the maximal fraction of trials in which Alice and Bob do not need to communicate trying to simulate a given behavior* $\{P_{ab|xy}\}_{xy}$*, optimized over all conceivable strategies with freely chosen settings.* **[♠]**

Similarly, we define a measure of free choice μ_F as

> *the maximal fraction of trials in which Alice and Bob can grant free choice of settings in trying to simulate a given behavior* $\{P_{ab|xy}\}_{xy}$*, optimized over all conceivable local strategies.* **[♣]**

In the quantum-mechanical context the measure μ_L is called a local fraction (27–32). By analogy, when considering the quantum-mechanical statistics the measure μ_F might be called a free fraction. This provides an equal basis for comparing the two assumptions within the standard causal (or realist) approach, which we formalize in the following section.

Causal Models, Locality, and Free Choice. The appropriate framework for the discussion of locality and free choice is provided by hidden-variable models (1–5). First, a hidden-variable model allows a formal statement of the realism assumption, understood to mean that properties of a physical system exist irrespective of an act of measurement (counterfactual definiteness). Second, hidden-variable models provide the causal language in which the locality and free choice assumptions are expressed (6, 7). The locality assumption conveys the requirement that the propagation of physical (or causal) influences have to follow the spatiotemporal structure of events (i.e., preserve the arrow of time and respect that actions at a distance require time). The free-choice assumption concerns the choice of measurement settings which are deemed causally unaffected by anything in the past (and thus it is sometimes called measurement independence).[‡] Both assumptions take the form of conditional independencies between certain variables in a hidden-variables model.

To make this idea more concrete, let us consider a given set of probability distributions (behavior) $\{P_{ab|xy}\}_{xy}$ which describes the statistics in a Bell experiment. Without loss of generality, by conditioning on λ in some a priori unknown hidden-variable space Λ, one can always write (4, 5, 7)

$$P_{ab|xy} = \sum_{\lambda\in\Lambda} P_{ab|xy\lambda} \cdot P_{\lambda|xy}, \qquad \textbf{[6]}$$

where $P_{\lambda|xy}$ and $P_{ab|xy\lambda}$ are valid (i.e., normalized) conditional probability distributions. The role of the hidden variable (cause in the past) $\lambda\in\Lambda$, distributed according to some P_λ, is to provide an explanation of the observed experimental statistics. This means that at each run of the experiment the outcomes are described by the distribution $P_{ab|xy\lambda}$ with $\lambda\in\Lambda$ fixed in a given trial, so that the accumulated experimental statistics $P_{ab|xy}$ obtains by sampling from some distribution $P_{\lambda|xy}$ over the whole hidden-variable space Λ. It is customary to say that

> *the choice of space* Λ *and probability distribution* P_λ *along with conditional distributions* $P_{ab|xy\lambda}$ *and* $P_{\lambda|xy}$ *satisfying Eq.* **6** *specify a hidden-variable (HV) model of a given behavior* $\{P_{ab|xy}\}_{xy}$*.* **[#]**

Note that such a model implicitly describes the distribution of settings chosen by Alice and Bob through the standard formula

$$P_{xy} = \sum_{\lambda\in\Lambda} P_{xy|\lambda} \cdot P_\lambda. \qquad \textbf{[7]}$$

So far the framework is general enough to accommodate any causal explanation of the statistics observed in the experiment. The assumptions of locality and free choice take the form of constraints on conditional distributions in **[#]**. For a local hidden-variable (LHV) model, we require the following factorization[§]:

$$P_{ab|xy\lambda} = P_{a|x\lambda} \cdot P_{b|y\lambda}, \qquad \textbf{[8]}$$

for each $x, y \in \mathfrak{M}$ and all $\lambda\in\Lambda$. The freedom of choice assumption consists of requiring that λ does not contain any information about variables x, y representing Alice and Bob's choice of measurement settings. This boils down to the independence condition (6, 22)

$$P_{\lambda|xy} = P_\lambda \qquad (\text{or equivalently } P_{xy|\lambda} = P_{xy}), \qquad \textbf{[9]}$$

holding for $x, y \in \mathfrak{M}$ and all $\lambda\in\Lambda$. In the following, we will abbreviate a hidden-variable model with freedom of choice as an FHV model.

The crucial point is the distinction between local vs. nonlocal as well as free vs. nonfree situations in the individual runs of the experiment modeled by Eq. **6**. This means that each condition Eqs. **8** and **9** should be considered separately for each $\lambda\in\Lambda$, i.e., whenever the respective condition does not hold for a given λ the assumption fails on the corresponding experimental trials. Such a

[†]For the simulation of a given behavior $\{P_{ab|xy}\}_{xy}$ in a Bell experiment one may proceed as follows. Upon rejection of locality, in each trial the system on Alice's side, one may not only use input x but also y to generate outcomes (and similarly for the box on Bob's side) that comply with the appropriate distribution. On the other hand, when freedom of choice is abandoned, both settings x, y may be specified in advance on each trial and the boxes can be instructed to provide the outcomes needed to simulate the appropriate distribution. It is, however, unclear how this might work with occasional violation of the respective assumptions.

[‡]As noted, the free-choice assumption is sometimes called measurement independence. Instead of on the agent, measurement independence is focused on the measurement devices and possible correlations between their settings, which can affect the observed statistics. Regardless of interpretation, the mathematics remains the same, with the source of correlations traced to some common factor (in the causal past).

[§]Locality can be seen as a conjunction of two conditions: parameter independence $P_{a|xy\lambda} = P_{a|x\lambda}$ and $P_{b|xy\lambda} = P_{b|y\lambda}$, and outcome independence $P_{a|bxy\lambda} = P_{a|xy\lambda}$ and $P_{b|axy\lambda} = P_{b|xy\lambda}$. One can show that such defined locality entails the factorization condition $P_{ab|xy\lambda} = P_{a|x\lambda} \cdot P_{b|y\lambda}$ (46).

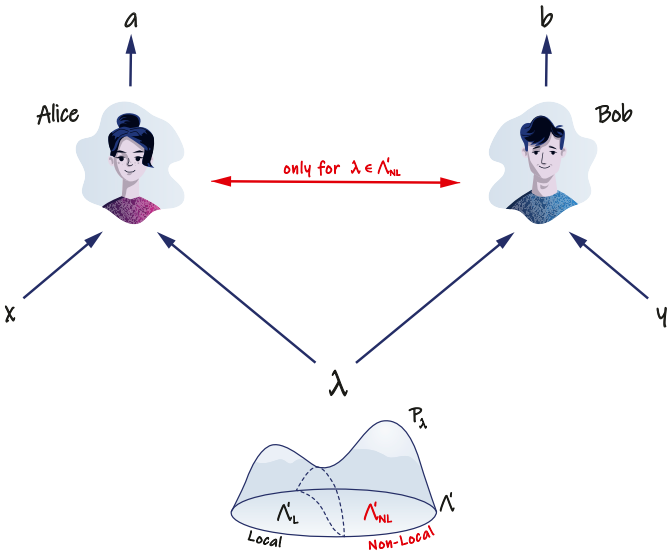


Fig. 2. Causal model with some nonlocality (communication). In a Bell scenario, with free choice of settings, correlations between Alice and Bob's outcomes have two possible explanations: common cause in the past or causal influence between the parties. In any causal model the space of hidden variables (representing common causes) splits into two disjoint parts $\Lambda' = \Lambda'_L \cup \Lambda'_{NL}$ distinguished by whether, for a given $\lambda \in \Lambda'$, causal influence occurs or not (Eq. **10**). Then, locality is measured by the proportion of events when locality is maintained, which is equal to the probability accumulated over subset Λ'_L, i.e., *Prob* $(\lambda \in \Lambda'_L) \equiv \sum_{\lambda \in \Lambda'_L} P_\lambda$.

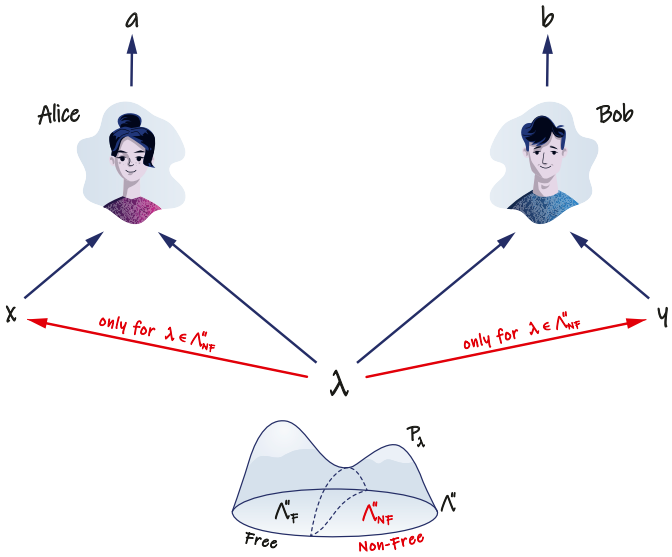


Fig. 3. Causal model with some freedom of choice (rigging). In a Bell scenario, with locality assumption, correlations between the outcomes on Alice and Bob's side can be explained by a common cause affecting choice or not (the latter implies freedom of choice). In any causal model the space of hidden variables (representing common causes) splits into two disjoint parts $\Lambda'' = \Lambda''_F \cup \Lambda''_{NF}$ distinguished by whether, for a given $\lambda \in \Lambda''$, the choice is free or not (Eq. **11**). Then, the parties enjoy freedom of choice only on the trials when $\lambda \in \Lambda''_F$, which happens with a frequency equal to the probability accumulated over subset Λ''_F, i.e., *Prob* $(\lambda \in \Lambda''_F) \equiv \sum_{\lambda \in \Lambda''_F} P_\lambda$.

distinction leads to a natural splitting of the underlying HV space into two unique partitions $\Lambda = \Lambda_L \cup \Lambda_{NL}$ and $\Lambda = \Lambda_F \cup \Lambda_{NF}$. The first one divides Λ by the locality property

$$\begin{aligned} \lambda \in \Lambda_L &\Leftrightarrow \text{Eq. } \mathbf{8} \text{ holds for } \textit{all } x, y \in \mathfrak{M}, \\ \lambda \in \Lambda_{NL} &\Leftrightarrow \text{Eq. } \mathbf{8} \text{ fails for } \textit{some } x, y \in \mathfrak{M}, \end{aligned} \qquad [10]$$

while the second one divides Λ by the free-choice property

$$\begin{aligned} \lambda \in \Lambda_F &\Leftrightarrow \text{Eq. } \mathbf{9} \text{ holds for } \textit{all } x, y \in \mathfrak{M}, \\ \lambda \in \Lambda_{NF} &\Leftrightarrow \text{Eq. } \mathbf{9} \text{ fails for } \textit{some } x, y \in \mathfrak{M}. \end{aligned} \qquad [11]$$

Figs. 2 and 3 illustrate the causal structures for two extreme cases: FHV and LHV models (in general built on different HV spaces Λ' and Λ''). The first one grants full freedom of choice ($\Lambda' = \Lambda'_F$) while allowing for partial violation of locality ($\Lambda' \supset \Lambda'_L$). The second one retains full locality ($\Lambda'' = \Lambda''_L$) while admitting some violation of free choice ($\Lambda'' \supset \Lambda''_F$).

Thus, for a given experimental trial (with $\lambda \in \Lambda$ fixed) the constraints in Eqs. **10** and **11** indicate, respectively, whether some nonlocal influence between the parties takes place ($\lambda \in \Lambda_{NL}$) and whether some influence from the past on the measurement settings occurs ($\lambda \in \Lambda_{NF}$). In other words, in a hypothetical simulation scenario these possibilities correspond to, respectively, communication or rigging measurement settings. How often this has to happen depends on the distribution P_λ. This picture lends itself to quantifying the degree of locality and freedom choice in a given HV model.

Remark 1. *For a given HV model [#] locality is measured by Prob* $(\lambda \in \Lambda_L) \equiv \sum_{\lambda \in \Lambda_L} P_\lambda$, *and similarly freedom of choice is measured by Prob* $(\lambda \in \Lambda_F) \equiv \sum_{\lambda \in \Lambda_F} P_\lambda$.

This remark captures the intuition of measuring locality and freedom of choice by considering the proportion of trials when the respective property is maintained across the whole experimental ensemble. We note that this quantity is model-dependent, since it is a property of a particular HV model adopted to explain some given experimental statistics $\{P_{ab|xy}\}_{xy}$ (including the distribution of measurement settings P_{xy}; cf. Eq. **7**).

The concepts just introduced allow a precise expression for the informal definitions **[♠]** and **[♣]** given above.

Definition 1. *For a given behavior* $\{P_{ab|xy}\}_{xy}$ *the measure of locality* μ_L *and freedom of choice* μ_F *are defined as*

$$\mu_L := \min_{P_{xy}} \max_{FHV} \sum_{\lambda \in \Lambda_L} P_\lambda, \qquad [12]$$

$$\mu_F := \min_{P_{xy}} \max_{LHV} \sum_{\lambda \in \Lambda_F} P_\lambda, \qquad [13]$$

where the maxima are taken respectively over all hidden-variable models with freedom of choice (FHV) or all local hidden-variable models (LHV) simulating given behavior $\{P_{ab|xy}\}_{xy}$, *with a fixed distribution of settings* P_{xy}, *minimized over any choice of the latter.*

This definition follows the intuition of, respectively, locality or free choice as properties that can be relaxed only to the extent that is required to maintain the other assumption in every experimental situation (i.e., for any distribution of measurement settings P_{xy}). Formally, the measures μ_L and μ_F count the maximal frequency of, respectively, local or free-choice events optimized over all protocols simulating $\{P_{ab|xy}\}_{xy}$ without violating of the other assumption (cf. *Remark 1*). The minimum over all P_{xy} amounts to the worst-case scenario, which takes into account the possibility that P_{xy} is a priori unspecified (i.e., this amount of freedom is enough to simulate an experiment with any arbitrary choice of distribution P_{xy} in compliance with Eq. **7**).

At first glance, even if conceptually appropriate, such a definition might seem too general to provide a manageable notion, due to the range of experimental scenarios that need to be taken into account (i.e., arbitrariness of P_{xy}). However, the situation

considerably simplifies because of the following lemma (see *Materials and Methods* for further discussion and proof). This lemma also provides additional support for *Definition 1*.

Lemma 1. *In both* Eqs. **12** and **13** *in Definition 1 the first minimum can be omitted, i.e., we have*

$$\mu_L = \max_{FHV} \sum_{\lambda\in\Lambda_L} P_\lambda, \qquad [14]$$

$$\mu_F = \max_{LHV} \sum_{\lambda\in\Lambda_F} P_\lambda, \qquad [15]$$

where the respective maxima are taken for some fixed nontrivial distribution P_{xy} (i.e., the expression is insensitive to this choice provided all settings are probed, $P_{xy}\neq 0$ for all x, y).

It is in this way that the present measure of locality μ_L extends the notion of local fraction (27–32) to arbitrary experimental behavior $\{P_{ab|xy}\}_{xy}$. Remarkably, the twin concept, which is the measure of free choice μ_F, has not been considered at all. Perhaps the reason for this omission is the issue of arbitrariness of the distribution P_{xy}, for which there are nontrivial constraints when freedom of choice is violated (note that for the measure μ_L this problem does not occur). Those concerns can be dismissed only after the proper treatment in *Lemma 1*. This allows a so-defined measure of free choice μ_F on a par with the more familiar measure of locality μ_L.

So far the concepts of violation of locality and freedom of choice, and the corresponding measures μ_L and μ_F, have been kept separate. This is expected given their disparate character. First, each concept plays a different role in the description of an experiment and hence offers a different explanation for any observed correlations, this is, direct influence (communication during the experiment) vs. measurement dependence (employing common past for rigging measurement settings). Second, on the level of causal modeling those assumptions are expressed differently, Eq. **8** vs. Eq. **9**. Third, violating free choice gives rise to subtle issues regarding constraints on the distribution of settings P_{xy} (as noted, these concerns are addressed in *Lemma 1*).

Having brought all those issues to the spotlight, it is surprising that the assumption of locality and free choice are intrinsically connected. We now present the key result in this paper showing the exchangeability of both concepts, while maintaining the same degree of locality and freedom of choice so defined. It holds for any number of settings $x, y \in \mathfrak{M} = \{1, \ldots, M\}$ (see *Materials and Methods* for the proof).

Theorem 1. *For a given behavior $\{P_{ab|xy}\}_{xy}$ the degree of locality and freedom of choice are the same, i.e., both measures in Definition 1 coincide $\mu_L = \mu_F$.*

This is a general structural theorem about causal modeling of a given behavior $\{P_{ab|xy}\}_{xy}$. It means that the resources measured by the frequency of causal interventions of either sort, required to explain an experimental statistics, are equally costly. Thus, as far as the statistics is concerned, causal explanations resorting either to violation of locality or free choice (or measurement dependence) should be kept on an equal footing. Preference should be guided by a better understanding of a particular situation (design of the experiment as well as ontological commitments in its description).

Let us emphasize two features of *Theorem 1*. First, this is a theory-independent result in the sense that it applies directly to experimental statistics irrespective of the design or theoretical framework behind the experiment (with the quantum predictions being just one example). Second, the connection between those two seemingly disparate quantities μ_L and μ_F has a practical advantage: Knowledge of one suffices to compute the other. Both features are illustrated by the following results.

Nonsignaling Behavior with Binary Settings. Consider the case of Bell's experiment with only two measurement settings on each side $x, y \in \mathfrak{M} = \{0, 1\}$. Let us recall that nonsignaling of some given behavior $\{P_{ab|xy}\}_{xy}$ means that Alice cannot infer Bob's measurement setting (whether it is $y = 0$ or 1) from the statistics on her side alone, i.e.,

$$P_{a|x0} = \sum_b P_{ab|x0} = \sum_b P_{ab|x1} = P_{a|x1} \quad \text{for all } a, x, \qquad [16]$$

and similarly on Bob's side (whether Alice chooses $x = 0$ or 1), i.e.,

$$P_{b|0y} = \sum_a P_{ab|0y} = \sum_a P_{ab|1y} = P_{b|1y} \quad \text{for all } b, y. \qquad [17]$$

Now, we can state another result which explicitly computes both measures μ_L and μ_F in a surprisingly simple form (see *Materials and Methods* for the proof).

Theorem 2. *For a given nonsignaling behavior $\{P_{ab|xy}\}_{xy}$ with binary settings $x, y \in \mathfrak{M} = \{0, 1\}$ both measures of locality μ_L and free choice μ_F from Definition 1 are equal to*

$$\mu_L = \mu_F = \begin{cases} \frac{1}{2}(4 - S_{max}), & \text{if } S_{max} > 2, \\ 1, & \text{otherwise,} \end{cases} \qquad [18]$$

where $S_{max} = \max\{|S_i| : i = 1, \ldots, 4\}$ is the maximum absolute value of the four CHSH expressions in Eqs. **2–5**.

We thus obtain a systematic method for assessing the degree of locality and free choice directly from the observed statistics $\{P_{ab|xy}\}_{xy}$ without reference to the specifics of the experiment (the only requirement is nonsignaling of the observed distributions). In this sense, this is a general theory-independent statement.

Overall, *Theorem 2* allows an interpretation of the amount of violation of the CHSH inequalities in Bell-type experiments as a fraction of trials violating locality (granted freedom of choice) or equivalently trials without freedom of choice (given locality).

The Quantum Case: Binary Settings and Beyond. Let us restrict our attention to the special case of the quantum statistics. Notably, various aspects of nonlocality have been extensively researched in relation to the quantum-mechanical predictions; see refs. 4 and 5 for a review. This includes the notion of local fraction (27–32), which is the same as measure μ_L here defined for a general behavior $\{P_{ab|xy}\}_{xy}$. As noted, it may be thus surprising that the equally natural measure of free choice μ_F has not been explored. *Theorem 1* bridges the gap between those two seemingly disparate notions: There is no actual need for separate study. We next review some crucial results for the local fraction in the quantum-mechanical framework, which allows us to make similar statements for the measure of free choice μ_F.

We first observe that *Theorem 2* can be readily applied to the quantum-mechanical statistics (where nonsignaling holds). In a Bell experiment, quantum probabilities obtain through the standard formula $P_{ab|xy} = Tr\,[\,\rho\,\mathbb{P}_x^a \otimes \mathbb{P}_y^b\,]$, where ρ is a (bipartite) mixed state with two projection-valued measures $\{\mathbb{P}_x^{a=\pm 1}\}$ and $\{\mathbb{P}_y^{b=\pm 1}\}$ representing Alice and Bob's choice of measurement settings $x, y \in \mathfrak{M} = \{0, 1\}$. Calculating the CHSH expressions Eqs. **2** and **5** in each particular case is straightforward, which gives explicitly the expression for both measures μ_L and μ_F via

PHYSICS

Eq. **18**. The result of special significance concerns the famous Tsirelson bound $S_{max}^{QM} = 2\sqrt{2}$ for the maximal violation of the CHSH inequalities in quantum mechanics (47). By virtue of *Theorem 2*, this means that in order to locally recover the quantum predictions in a Bell experiment with two settings Alice and Bob can enjoy freedom of choice in the worst case, at most, with a fraction $\mu_F = 2 - \sqrt{2} \approx 0.59$ of all trials (corresponding to the choice of measurements on a maximally entangled state that saturate the Tsirelson bound). Clearly, the same applies to local fraction μ_L in a two-setting scenario.

Interestingly, relaxing the constraint on the number of settings for Alice and Bob's measurements $x, y \in \mathfrak{M} = \{1, 2, 3, \ldots, M\}$ the quantum statistics forces us to further constrain, respectively, locality or free choice. The case of local fraction μ_L with arbitrary number of settings $M \to \infty$ has been thoroughly investigated for statistics generated by quantum states. Let us refer to two interesting results in the literature on local fraction μ_L which readily translate via *Theorem 1* to the measure of free choice μ_F. The first one concerns the statistics of a maximally entangled state (cf. refs. 27 and 29) (see *SI Appendix* for a direct proof).

Theorem 3. *For every LHV model that explains the statistics of a Bell experiment for a maximally entangled state the amount of free choice tends to zero with increasing number of measurement settings* M, *i.e.*, $\mu_F \underset{M \to \infty}{\longrightarrow} 0$.

Apparently, for less entangled states the amount of freedom increases, reaching the maximal value $\mu_F = 1$ for separable states. This is a consequence of the result in ref. 32, which explicitly computes the local fraction μ_L for all pure two-qubit states. Stated for measure μ_F this takes the following form.

Theorem 4. *For a pure two-qubit state, which by appropriate choice of the basis can always be written in the form* $|\psi\rangle = \cos\frac{\theta}{2}|00\rangle + \sin\frac{\theta}{2}|11\rangle$ *with* $\theta \in [0, \frac{\pi}{2}]$, *the amount of free choice is equal* $\mu_F = \cos\theta$, *whatever the choice and number of settings on Alice and Bob's side*.

Note that both *Theorem 3* and *Theorem 4* assume a specific form of behavior $\{P_{ab|xy}\}_{xy}$ as obtained by the rules of quantum theory. The theorems should be contrasted with *Theorem 2*, which is a theory-independent statement, not limited to a particular theoretical framework.

Discussion

The ingenuity of Bell's theorem lies in the fundamental nature of the premises from which the result is derived. Within the standard causal (or realist) approach, it is hard to assume less about two agents than having free choice and their systems being localized in space. Yet, in some experiments nature refutes the possibility that both assumptions are concurrently true (8–15). It is not easy to reject either one of them without carefully rethinking the role of observers and how cause-and-effect manifests in the world.[¶] Our objective in this paper is this: *Instead of pondering the question of how this could be possible, we ask about the extent to which a given assumption has to be relaxed in order to maintain the other.* Expressed more colloquially, it is natural for a realist to ask what the cost is of trading one concept for the other: *Is it possible to save free choice by giving up on some locality, or maybe is it better to forego a modicum of free choice in exchange for locality?* These questions can be compared on equal footing by computing a proportion of trials across the whole experimental ensemble in which a given assumption must fail, when the other holds at all times. Surprisingly, the answer can be obtained by looking at the observed statistics alone (avoiding the specifics of the experimental setup). The first question was formulated in the quantum-mechanical context by Elitzur et al. (27), who introduced the notion of local fraction further elaborated in refs. 29–32 (see ref. 28 for an early indication of these ideas). Here, we generalize this notion to arbitrary experimental statistics (see also ref. 33). Furthermore, we answer the second question by adopting a similar approach to measuring the amount of free choice (which by analogy may be called free fraction). The first main result, *Theorem 1*, compares such defined measures in the general case (arbitrary statistics with any number of settings), showing that both assumptions are equally costly. This demonstrates a deeper symmetry between locality and free choice, which may come as a surprise, given our intuition of a profound difference in the role these concepts play in the description of an experiment.

In this paper, the notions of locality and free choice are understood in the usual sense required to derive Bell's theorem (6, 22). They are expressed in the standard causal model framework (which subsumes realism) as unambiguous yes–no criteria for each experimental trial (i.e., when all past variables are fixed), determining whether there is a causal link between certain variables in a model (without pondering its exact nature). The measures μ_L and μ_F count the fraction of trials when such a connection needs to be established, breaking locality or free choice, respectively, in order to explain the observed statistics. This problem is prior to a discussion of how this actually occurs, which is particularly relevant when the exact nature of the phenomenon under study is obscured. *Theorem 1* shows no intrinsic reason for a realist to favor one assumption vs. the other. The minimal frequency of the required causal influences of either sort, measured by μ_L and μ_F, is exactly the same. Notably, this is a general result which holds for any behavior $\{P_{ab|xy}\}_{xy}$. What remains is explicit calculation of those measures for a given experimental statistics.

The second main result, *Theorem 2*, evaluates both measures μ_L and μ_F for any nonsignaling behavior in a Bell experiment with two outcomes and two settings. It provides a direct interpretation to the amount of violation of the CHSH inequalities (43). The key motivation behind this result is that the degree by which the inequalities are violated has not been given tangible interpretation so far, beyond its use as a binary test of whether the inequalities are obeyed or not in study of Bell nonlocality. Furthermore, *Theorem 2* has the advantage of being theory-independent in the sense of being applicable to the experimental statistics regardless of its theoretical origin (i.e., beyond the quantum-mechanical framework). This makes it suitable for quantitative assessment of the degree of locality and free choice across different experimental situations, with prospective applications beyond physics, e.g., in neuroscience, cognitive psychology, social sciences, or finance (48–52).

We also state two results, *Theorem 3* and *Theorem 4*, for the measure of free choice μ_F in the case of the quantum statistics generated by the pure two-qubit states. Both are direct translation, via *Theorem 1*, of the corresponding results for the local fraction μ_L (27–32).

It is worth noting a related idea of quantifying nonlocality through the amount of information transmitted between the parties that is required to reproduce quantum correlations (under free-choice assumption). Together with the development of the specific models (53–57), this has led to various results regarding communication complexity in the quantum realm (58). However, in this paper we take a different perspective on measuring nonlocality by changing the question from "how much" to "how often" communication needs to be established between the

[¶]We note that the conventional understanding of causality and the language of counterfactuals has recently gained a solid mathematical basis; see e.g., the work of Pearl (7). However, in view of the apparent difficulties with embedding quantum mechanics in that framework, the standard approach to causality based on Reichenbach's principle or claims regarding spatiotemporal structure of events might need reassessment; see, e.g., indefinite causal structures (23, 24) or retrocausality (25).

parties to simulate given correlations. *Theorem 2* gives the exact bound in the case of nonsignaling statistics in the two-setting and two-outcome Bell experiment. In the quantum case, such a simulation requires communication in at least 41% of trials [because of Tsirelson's bound (47)] and for maximally entangled states increases to 100% of trials when the number of settings is arbitrary (cf. *Theorems 3* and *4*).

Natural as it may seem, the idea of measuring freedom of choice by measure μ_F has not been developed in the literature. The reason for this omission can be traced to the conceptual and technical issues with handling arbitrariness of the distribution of settings P_{xy}. Those concerns are properly addressed in the present paper with *Lemma 1*, which considerably simplifies and supports *Definition 1*. We note that various measures have been developed as a means of quantifying freedom of choice (or measurement independence, as it is sometimes called). They include maximal distance between distributions (35, 37), mutual information (38, 42), or measurement-dependent locality (39–41). Furthermore, some explicit models simulating correlations in a singlet state with various degrees of measurement dependence have been proposed (34, 36) and analyzed (e.g., see ref. 42 for comparison of causal vs. retrocausal models). However, these attempts depart from the original understanding of the free choice as introduced by Bell (6, 22) (i.e., strict independence of choice from anything in the past) in favor of more sophisticated information-theoretic accounts. Notably, the proposed measure of free choice builds on Bell's original framework assessing the maximal frequency with which such a freedom can be retained in a model strictly consistent with locality. It thus benefits from a direct interpretation within the established causal framework of Bell inequalities and has a clear-cut operational meaning.

Regarding *Theorem 3*, which rules out any freedom of choice so defined, it is interesting to take an adversarial perspective on the problem of free choice in relation to quantum cryptography and device-independent certification (59, 60). In this narrative an eavesdropper controls the devices trying to simulate the quantum statistics of a Bell test, which is impossible as long as the parties enjoy freedom of choice. However, any breach of the latter, i.e., control of measurement settings, shifts the balance in favor of the eavesdropper in her malicious task. Taking the view that any causal influence comes with a cost or danger of being uncovered there are two diverging strategies that reduce the cost/risk to be considered: 1) resort to the use of control of choice as seldom as possible during the experiment or 2) minimize the intensity of each act of control. *Theorem 3* completely rules out the first possibility when simulating quantum statistics, i.e., the eavesdropper needs to manipulate both settings on each trial in order to simulate the quantum predictions. The question about the intensity of the control is left open in our discussion but amenable to information-theoretic methods (35–42). This gives additional security criteria for quantum cryptography and device-independent certification by forcing the eavesdropper to a more challenging sort of attack (not only can she not miss a trial but also the control has to be subtle enough).

We remark that the main *Theorem 1* readily extends to the case of larger number of parties and outcomes $\{P_{abc\ldots|xyz\ldots}\}_{xyz\ldots}$. This should be also possible for *Theorem 2* when characterization of the local polytope is known (cf. refs. 61–67). Yet another valuable avenue for research in that case consists of completing the analysis to include signaling scenarios (68, 69). As for the quantum case, we considered the simplest Bell-type scenario with two parties involved in the experiment, but extensions may prove even more surprising (see ref. 5 for a technical review of the vast field of Bell nonlocality). In particular, in three-party scenarios the methods discussed presently can be used to eliminate freedom of choice already for two settings per party sharing the Greenberger–Horne–Zeilinger state [cf. Mermin inequalities which saturate in that case (70)]. We should also mention an intriguing result (71) for a triangle quantum network in which nonlocality can be proved with all measurements fixed. Remarkably, there is nothing to choose in that setup, but there is another assumption of preparation independence which plays a crucial role in the argument.

In this paper we are trying to remain impartial as to which assumption—locality or free choice—is more important on the fundamental level. This is certainly a strongly debated subject in general, both among physicists and philosophers, with strong supporters on each side (16–18). As just one example depreciating the role of freedom of choice let us quote Albert Einstein[‖]: "Human beings, in their thinking, feeling and acting are not free agents but are as causally bound as the stars in their motion." As a counterbalance, it is hard to resist the objection that was eloquently stated by Nicolas Gisin (ref. 72, p. 90): "But for me, the situation is very clear: not only does free will exist, but it is a prerequisite for science, philosophy, and our very ability to think rationally in a meaningful way." Without entering into this debate, we remark that both assumptions are interchangeable on a deeper level. Namely, for a given experimental statistics $\{P_{ab|xy}\}_{xy}$ in a Bell-type experiment the measure of locality μ_L and measure of free choice μ_F are exactly the same. This makes an even stronger case regarding the inherent impossibility of inferring causal structure from experimental statistics alone.

Materials and Methods

In order to facilitate the following discussion we begin with two technical lemmas. See *SI Appendix* for the proofs.

The first one holds for a Bell experiment with arbitrary number of settings $x, y \in \mathfrak{M} = \{1, 2, 3, \ldots, M\}$.

Lemma 2. *For any behavior $\{P_{ab|xy}\}_{xy}$ and distribution of settings P_{xy} there exists a local hidden-variable model (LHV) which fully violates the freedom of choice assumption (i.e., if $\tilde{\Lambda}$ is the relevant HV space, then we have $\tilde{\Lambda} = \tilde{\Lambda}_L = \tilde{\Lambda}_{NF}$; cf.* Eqs. **10** and **11**).

The second one concerns a Bell scenario with binary settings $x, y \in \mathfrak{M} = \{0, 1\}$.

Lemma 3. *Each nonsignaling behavior $\{P_{ab|xy}\}_{xy}$ with binary settings $x, y \in \mathfrak{M} = \{0, 1\}$ can be decomposed as a convex mixture of a local behavior $\{\bar{P}_{ab|xy}\}_{xy}$ and a PR-box $\{\tilde{P}_{ab|xy}\}_{xy}$ in the form*

$$P_{ab|xy} = p \cdot \bar{P}_{ab|xy} + (1-p) \cdot \tilde{P}_{ab|xy}, \qquad [19]$$

with $p = \frac{1}{2}(4 - S_{max})$ for all $x, y \in \{0, 1\}$.

Recall that a PR-box (73) is a nonsignaling behavior for which one of the CHSH expressions in Eqs. **2**–**5** reaches the maximal algebraic bound of $|S_i| = 4$. Here, local behavior means existence of an LHV + FHV model of $\{\bar{P}_{ab|xy}\}_{xy}$ and $S_{max} = \max\{|S_i| : i = 1, \ldots, 4\}$.

We are now ready to proceed with the proofs.

Proof of Lemma 1: Suppose we have an HV model [#] of some behavior $\{P_{ab|xy}\}_{xy}$ for some nontrivial distribution of settings P_{xy}. The latter obtains via Eq. **7** from the conditional probabilities $P_{xy|\lambda}$ which are related to probabilities specified by the model, $P_{\lambda|xy}$ and P_λ, by the usual Bayes' rule. The point at issue is whether a given HV model can simulate any other distribution of settings $\tilde{P}_{xy}$ via Eq. **7** by changing $P_{xy|\lambda} \rightsquigarrow \tilde{P}_{xy|\lambda}$, while keeping the remaining components of the HV model ($*$) intact. This requires consistency with Bayes' rule, i.e.,

$$\tilde{P}_{xy|\lambda} = \frac{P_{\lambda|xy} \cdot \tilde{P}_{xy}}{P_\lambda}, \qquad [20]$$

which should be a well-defined probability distribution for each λ. Since distributions $P_{\lambda|xy}$ and P_λ are fixed by the HV model [#], then the distribution of settings $\tilde{P}_{xy}$ is arbitrary as long as the expression in Eq. **20** is less than 1 for each $\lambda \in \Lambda$ (normalization is trivially fulfilled). Now, whenever freedom of choice from Eq. **9** holds, this condition is always satisfied, and

[‖] Statement to the Spinoza Society of America, 22 September 1932. AEA 33–291.

hence such an HV model can be trivially adjusted for any distribution $\tilde{P}_{xy}$ [by redefining $\tilde{P}_{xy|\lambda} := \tilde{P}_{xy}$ in compliance with Eq. **20** and keeping all of the remaining components of the HV model [#] unchanged]. Of course, for FHV models in the definition of μ_L in Eq. **12** this is the case, which thus entails the simpler expression for μ_L in Eq. **14**.

Clearly, such a simple argument falls apart for models without freedom of choice, like those in the definition of μ_F in Eq. **13**, when $P_{\lambda|xy}$ and P_λ do not cancel out and the probability in Eq. **20** may be ill-defined. In that case, some deeper intervention into the model is required as shown below.

Let us take some LHV model [#] simulating a given behavior $\{P_{ab|xy}\}_{xy}$ with nontrivial distribution of settings P_{xy}. Then, the related HV space decomposes as $\Lambda = \Lambda_F \uplus \Lambda_{NF}$ and the degree of freedom is measured by $p_F := \sum_{\lambda\in\Lambda_F} P_\lambda$ (cf. *Remark 1*). Now, consider a restriction of the model to the respective subspaces Λ_F and Λ_{NF} which amounts to the following rescaling:

$$P^F_\lambda := \tfrac{1}{p_F} P_\lambda\,, \quad P^F_{\lambda|xy} := \tfrac{1}{p_F} P_{\lambda|xy}\,, \quad P^F_{ab|xy\lambda} := P_{ab|xy\lambda}, \qquad [21]$$

for $\lambda \in \Lambda_F$, and similarly

$$P^{NF}_\lambda := \tfrac{1}{1-p_F} P_\lambda\,, \quad P^{NF}_{\lambda|xy} := \tfrac{1}{1-p_F} P_{\lambda|xy}\,, \quad P^{NF}_{ab|xy\lambda} := P_{ab|xy\lambda}, \qquad [22]$$

for $\lambda \in \Lambda_{NF}$. Both are LHV models with marginals

$$P^F_{ab|xy} = \sum_{\lambda\in\Lambda_F} P^F_{ab|xy\lambda} \cdot P^F_{\lambda|xy}, \qquad [23]$$

$$P^{NF}_{ab|xy} = \sum_{\lambda\in\Lambda_{NF}} P^{NF}_{ab|xy\lambda} \cdot P^{NF}_{\lambda|xy}, \qquad [24]$$

which provide a convex decomposition of the original behavior $\{P_{ab|xy}\}_{xy}$, i.e.,

$$P_{ab|xy} = p_F \cdot P^F_{ab|xy} + (1-p_F)\cdot P^{NF}_{ab|xy}. \qquad [25]$$

The crucial point is a careful adjustment of these two models to recover some arbitrary distribution of settings $\tilde{P}_{xy}$, while maintaining the respective marginals Eqs. **23** and **24**. For the first one (restriction to Λ_F) the situation is trivial as explained above: Since it is a FHV model, then it suffice to redefine $\tilde{P}^F_{xy|\lambda} := \tilde{P}_{xy}$ (in compliance with Eq. **20**) and leave all the rest intact. As for the second one (restriction to Λ_{NF}), we can use *Lemma 2* for constructing another HV space $\tilde{\Lambda}_{NF}$ with an LHV model without any free choice that simulates behavior $\{P^{NF}_{ab|xy}\}_{xy}$ with the required distribution of settings $\tilde{P}_{xy}$. Then, such modified models can be stitched back together on the compound HV space $\tilde{\Lambda} := \Lambda_F \uplus \tilde{\Lambda}_{NF}$ with respective weights p_F and $1-p_F$. This guarantees reconstruction of the original behavior $\{P_{ab|xy}\}_{xy}$ (see Eq. **25**) with the new distribution of settings $\tilde{P}_{xy}$. The model is local and has the same degree of freedom equal to p_F (the first component has full freedom of choice, while in the second one it is entirely missing).

The above construction shows that for every LHV model of some behavior $\{P_{ab|xy}\}_{xy}$ there is always another one adjusted for any other distribution of settings $\tilde{P}_{xy}$ with the same degree of freedom. This justifies the simpler expression for μ_F in Eq. **15** and hence concludes the proof of *Lemma 1*.

Proof of Theorem 1: Note that *Lemma 1* Eqs. **14** and **15** can be taken as a definition of measures μ_L and μ_F. This is very convenient, since it allows a discussion free from any concerns about the distribution of settings P_{xy} (this is particularly relevant in the case of μ_F as explained above).

It is instructive to observe that the calculation of both measures μ_L and μ_F can be succinctly formulated as a convex optimization problem. Suppose, we can decompose some given behavior $\{P_{ab|xy}\}_{xy}$ as a mixture

$$P_{ab|xy} = p_L \cdot P^L_{ab|xy} + (1-p_L)\cdot P^{NL}_{ab|xy}, \qquad [26]$$

where $\{P^L_{ab|xy}\}_{xy}$ is a local behavior with full freedom of choice (i.e., has an LHV + FHV model), and $\{P^{NL}_{ab|xy}\}_{xy}$ is a free behavior (i.e., has an FHV model). Also similarly, suppose that

$$P_{ab|xy} = p_F \cdot P^F_{ab|xy} + (1-p_F)\cdot P^{NF}_{ab|xy}, \qquad [27]$$

where $\{P^F_{ab|xy}\}_{xy}$ is a local behavior with full freedom of choice (i.e., has an LHV + FHV model) and $\{P^{NF}_{ab|xy}\}_{xy}$ is a local behavior (i.e., has an LHV model). In both cases we assume that $0 \leqslant p_L\,, p_F \leqslant 1$, and both Eqs. **26** and **27** have to hold for all $a, b = \pm 1$ and $x, y \in \mathfrak{M}$. Then, we have

Remark 2. *Measures μ_L and μ_F evaluate the maxima over all possible decompositions in* Eqs. **26** and **27** *of behavior $\{P_{ab|xy}\}_{xy}$, i.e.*,

$$\mu_L = \max_{\text{decomp. (26)}} p_L, \qquad [28]$$

$$\mu_F = \max_{\text{decomp. (27)}} p_F. \qquad [29]$$

Proof:
We will justify only Eq. **28**, since the argument for Eq. **29** is analogous.

Let us observe that every HV model [#] of behavior $\{P_{ab|xy}\}_{xy}$ as described by Eq. **6** splits into two components (cf. Eq. **10**):

$$P_{ab|xy} = \underbrace{\sum_{\lambda\in\Lambda_L} P_{ab|xy\lambda}\cdot P_\lambda}_{p_L\,\cdot\, P^L_{ab|xy}} + \underbrace{\sum_{\lambda\in\Lambda_{NL}} P_{ab|xy\lambda}\cdot P_{\lambda|xy}}_{(1-p_L)\,\cdot\, P^{NL}_{ab|xy}}, \qquad [30]$$

which defines decomposition of the type in Eq. **26** with $p_L := \sum_{\lambda\in\Lambda_L} P_\lambda$. Therefore, by Eq. **14**, we get $\mu_L \leqslant \max_{\text{decomp. (26)}} p_L$.

To see the reverse, we note that every decomposition of the type in Eq. **26** implies existence of an LHV + FHV model of behavior $\{P^L_{ab|xy}\}_{xy}$ on some HV space $\tilde{\Lambda}_L$ and a FHV model of behavior $\{P^{NL}_{ab|xy}\}_{xy}$ on some HV space $\tilde{\Lambda}_{NL}$. Those two models, when combined on a compound HV space $\Lambda := \tilde{\Lambda}_L \uplus \tilde{\Lambda}_{NL}$ with the respective weights p_L and $1-p_L$, provide an HV model of behavior $\{P_{ab|xy}\}_{xy}$. Since the local domain of such a model contains $\tilde{\Lambda}_L$, then from Eq. **14** we have $\mu_L \geqslant p_L$, which entails $\mu_L \geqslant \max_{\text{decomp. (26)}} p_L$. This concludes the proof of Eq. **28**. □

Now, in order to prove *Theorem 1* it is enough to show that for every decomposition of the type in Eq. **26** there exists a decomposition of the type in Eq. **27** with the same weight $p_L = p_F$, and vice versa. A closer look at both expressions reveals that behaviors $\{P^L_{ab|xy}\}_{xy}$ and $\{P^F_{ab|xy}\}_{xy}$ are both local with full freedom of choice (i.e., share the same LHV + FHV model). Thus, the problem can be reduced to justifying that 1) behavior $\{P^{NL}_{ab|xy}\}_{xy}$ also has an LHV model (possibly a non-FHV model) and 2) behavior $\{P^{NF}_{ab|xy}\}_{xy}$ also has an FHV model (possibly a non-LHV model).

Ad. 1: This readily follows from *Lemma 2*.

Ad. 2: Here, a trivial model will suffice. Let us take $\Lambda := \{\lambda_o\}$ (a single-element set) with $P_{\lambda_o} \equiv P_{\lambda_o|xy} := 1$ and conditional distribution defined as $P_{ab|xy\lambda_o} := P^{NF}_{ab|xy}$. Clearly, it is an FVH model of behavior $\{P^{NF}_{ab|xy}\}_{xy}$.

Thus, we have shown equivalence of both decompositions Eqs. **26** and **27**, which, by virtue of *Remark 2*, proves *Theorem 1*.

Proof of Theorem 2: By virtue of *Theorem 1* it suffices to prove the result for one of the measures. Let it be measure μ_L evaluated by Eq. **28** in *Remark 2*.

Consider some arbitrary decomposition Eq. **26** of behavior $\{P_{ab|xy}\}_{xy}$. Then, by linearity, the four CHSH expressions Eqs. **2**–**5** decompose as well, i.e., we get

$$S_i = p_L \cdot S^L_i + (1-p_L)\cdot S^{NL}_i, \qquad [31]$$

where S^L_i and S^{NL}_i are calculated for the respective behaviors $\{P^L_{ab|xy}\}_{xy}$ and $\{P^{NL}_{ab|xy}\}_{xy}$. Since the first one is a local behavior with full freedom of choice (i.e., having an LHV + FHV model), then from the CHSH inequalities Eq. **1** we have $|S^L_i| \leqslant 2$. For the second one there is nothing interesting to be said other than noting the maximal algebraic bound $|S^{NL}_i| \leqslant 4$. As a consequence, the following inequality obtains:

$$|S_i| \leqslant p_L \cdot 2 + (1-p_L)\cdot 4 = 4 - 2\,p_L, \qquad [32]$$

and we get $p_L \leqslant \frac{1}{2}(4-|S_i|)$. Thus, by assumed arbitrariness of decomposition, Eq. **26** gives the upper bound on expression in Eq. **28**:

$$\mu_L \leqslant \tfrac{1}{2}(4-|S_i|), \qquad [33]$$

where $S_{max} = \max\{|S_i| : i = 1, \ldots, 4\}$. By *Lemma 3* we conclude that the bound is tight, which ends the proof of *Theorem 2*.

Data Availability. There are no data underlying this work.

ACKNOWLEDGMENTS. We thank R. Colbeck, J. Duda, N. Gisin, M. Hall, L. Hardy, D. Kaiser, M. Markiewicz, S. Pironio, D. Rohrlich, and V. Scarani for helpful comments. P.B. acknowledges support from the Polish National Agency for Academic Exchange in the Bekker Scholarship Program. P.B. and E.M.P. were supported by Office of Naval Research Global grant N62909-19-1-2000.

1. J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics* (Cambridge University Press, 1987).
2. N. D. Mermin, Hidden variables and the two theorems of John Bell. *Rev. Mod. Phys.* **65**, 803–815 (1993).
3. H. M. Wiseman, The two Bell's theorems of John Bell. *J. Phys. A: Math. Theor.* **47**, 424001 (2014).
4. N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, S. Wehner, Bell nonlocality. *Rev. Mod. Phys.* **86**, 419 (2014).
5. V. Scarani, *Bell Nonlocality* (Oxford University Press, 2019).
6. J. S. Bell, *Free Variables and Local Causality in Speakable and Unspeakable in Quantum Mechanics* (Cambridge University Press, 1987).
7. J. Pearl, *Causality: Models, Reasoning, and Inference* (Cambridge University Press, ed. 2, 2009).
8. A. Aspect, J. Dalibard, G. Roger, Experimental test of Bell's inequalities using time-varying analyzers. *Phys. Rev. Lett.* **49**, 1804 (1982).
9. M. Giustina *et al.*, Significant-loophole-free test of Bell's theorem with entangled photons. *Phys. Rev. Lett.* **115**, 250401 (2015).
10. L. K. Shalm *et al.*, Strong loophole-free test of local realism. *Phys. Rev. Lett.* **115**, 250402 (2015).
11. B. Hensen *et al.*, Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres. *Nature* **526**, 682–686 (2015).
12. A. Aspect, Closing the door on Einstein and Bohr's quantum debate. *Physics* **8**, 123 (2015).
13. J. Gallicchio, A. S. Friedman, D. I. Kaiser, Testing Bell's inequality with cosmic photons: Closing the setting-independence loophole. *Phys. Rev. Lett.* **112**, 110405 (2014).
14. D. Rauch *et al.*, Cosmic Bell test using random measurement settings from high-redshift quasars. *Phys. Rev. Lett.* **121**, 080403 (2018).
15. C. Abellán *et al.*, Challenging local realism with human choices. *Nature* **557**, 212 (2018).
16. T. Maudlin, *Philosophy of Physics: Quantum Theory* (Princeton University Press, 2019).
17. F. Laloë, *Do We Really Understand Quantum Mechanics?* (Cambridge University Press, ed. 2, 2019).
18. T. Norsen, *Foundations of Quantum Mechanics: An Exploration of the Physical Meaning of Quantum Theory (Undergraduate Lecture Notes in Physics,* Springer, 2017).
19. C. J. Wood, 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).
20. R. Chaves, R. Kueng, J. B. Brask, D. Gross, Unifying framework for relaxations of the causal assumptions in Bell's theorem. *Phys. Rev. Lett.* **114**, 140403 (2015).
21. E. G. Cavalcanti, Classical causal models for Bell and Kochen-Specker inequality violations require fine-tuning. *Phys. Rev. X* **8**, 021018 (2018).
22. R. Colbeck, R. Renner, A short note on the concept of free choice. arXiv [Preprint] (2013). https://arxiv.org/abs/1302.4446 (Accessed 31 March 2021).
23. C. Brukner, Quantum causality. *Nat. Phys.* **10**, 259–263 (2014).
24. J. M. A. Allen, J. Barrett, D. Horsman, C. M. Lee, R. W. Spekkens, Quantum common causes and quantum causal models. *Phys. Rev. X* **7**, 031021 (2017).
25. K. B. Wharton, N. Argaman, Colloquium: Bell's theorem and locally mediated reformulations of quantum mechanics. *Rev. Mod. Phys.* **92**, 021002 (2020).
26. T. Norsen, Against 'realism'. *Found. Phys.* **37**, 311–340 (2007).
27. A. C. Elitzur, S. Popescu, D. Rohrlich, Quantum nonlocality for each pair in an ensemble. *Phys. Lett. A* **162**, 25–28 (1992).
28. L. Hardy, A new way to obtain Bell inequalities. *Phys. Lett. A* **161**, 21–25 (1991).
29. J. Barrett, A. Kent, S. Pironio, Maximally nonlocal and monogamous quantum correlations. *Phys. Rev. Lett.* **97**, 170409 (2006).
30. R. Colbeck, R. Renner, Hidden variable models for quantum theory cannot have any local part. *Phys. Rev. Lett.* **101**, 050403 (2008).
31. R. Colbeck, R. Renner, "The completeness of quantum theory for predicting measurement outcomes" in *Quantum Theory: Informational Foundations and Foils*, G. Chiribella, R. W. Spekkens, Eds. (Springer, 2016), pp. 497–528.
32. S. Portmann, C. Branciard, N. Gisin, Local content of all pure two-qubit states. *Phys. Rev. A* **86**, 012104 (2012).
33. S. Abramsky, R. S. Barbosa, S. Mansfield, Contextual fraction as a measure of contextuality. *Phys. Rev. Lett.* **119**, 050504 (2017).
34. C. H. Brans, Bell's theorem does not eliminate fully causal hidden variables. *Int. J. Theor. Phys.* **27**, 219–226 (1988).
35. M. J. W. Hall, Local deterministic model of singlet state correlations based on relaxing measurement independence. *Phys. Rev. Lett.* **105**, 250404 (2010).
36. M. J. W. Hall, "The significance of measurement independence for Bell inequalities and locality" in *At the Frontier of Spacetime*, T. Asselmeyer-Maluga, Ed. (Springer, 2016), pp. 189–204.
37. M. J. W. Hall, Relaxed Bell inequalities and Kochen-Specker theorems. *Phys. Rev. A* **84**, 022102 (2011).
38. J. Barrett, N. Gisin, How much measurement independence is needed to demonstrate nonlocality? *Phys. Rev. Lett.* **106**, 100406 (2011).
39. G. Pütz, D. Rosset, T. J. Barnea, Y. C. Liang, N. Gisin, Arbitrarily small amount of measurement independence is sufficient to manifest quantum nonlocality. *Phys. Rev. Lett.* **113**, 190402 (2014).
40. D. Aktas *et al.*, Demonstration of quantum nonlocality in the presence of measurement dependence. *Phys. Rev. Lett.* **114**, 220404 (2015).
41. G. Pütz, N. Gisin, Measurement dependent locality. *New J. Phys.* **18**, 055006 (2016).
42. M. J. W. Hall, C. Branciard, Measurement-dependence cost for Bell nonlocality: Causal versus retrocausal models. *Phys. Rev. A* **102**, 052228 (2020).
43. J. F. Clauser, M. A. Horne, A. Shimony, R. A. Holt, Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.* **23**, 880–884 (1969).
44. A. Fine, Hidden variables, joint probability, and the Bell inequalities. *Phys. Rev. Lett.* **48**, 291–295 (1982).
45. J. J. Halliwell, Two proofs of Fine's theorem. *Phys. Lett. A* **378**, 2945–2950 (2014).
46. J. P. Jarrett, On the physical significance of the locality conditions in the Bell arguments. *Noûs* **18**, 569 (1984).
47. B. S. Tsirelson, Quantum generalizations of Bell's inequality. *Lett. Math. Phys.* **4**, 93–100 (1980).
48. J. R. Busemeyer, P. D. Bruza, *Quantum Models of Cognition and Decision* (Cambridge University Press, 2012).
49. E. Haven, A. Khrennikov, *Quantum Social Science* (Cambridge University Press, 2013).
50. E. M. Pothos, J. R. Busemeyer, Can quantum probability provide a new direction for cognitive modeling? *Behav. Brain Sci.* **36**, 255–274 (2013).
51. L. Hardy, "Bell inequalities with retarded settings" in *Quantum [Un]speakables II: Half a Century of Bell's Theorem*, R. Bertlmann, A. Zeilinger, Eds. (Springer, 2017), pp. 261–272.
52. L. Hardy, Proposal to use humans to switch settings in a bell experiment. arXiv [Preprint] (2017). https://arxiv.org/abs/1705.04620 (Accessed 31 March 2021).
53. T. Maudlin, "Bell's inequality, information transmission, and prism models" in *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association 1992* (The University of Chicago Press, 1992), vol. 1, pp. 404–417.
54. G. Brassard, R. Cleve, A. Tapp, Cost of exactly simulating quantum entanglement with classical communication. *Phys. Rev. Lett.* **83**, 1874 (1999).
55. M. Steiner, Towards quantifying non-local information transfer: Finite-bit non-locality. *Phys. Lett. A* **270**, 239 (2000).
56. B. F. Toner, D. Bacon, Communication cost of simulating Bell correlations. *Phys. Rev. Lett.* **91**, 187904 (2003).
57. R. D. Gill, The triangle wave versus the cosine: How classical systems can optimally approximate EPR-B correlations. *Entropy* **22**, 287 (2020).
58. H. Buhrman, R. Cleve, S. Massar, R. de Wolf, Nonlocality and communication complexity. *Rev. Mod. Phys.* **82**, 665–698 (2010).
59. J. Kofler, T. Paterek, C. Brukner, Experimenter's freedom in Bell's theorem and quantum cryptography. *Phys. Rev. A* **73**, 022104 (2006).
60. D. E. Koh *et al.*, Effects of reduced measurement independence on Bell-based randomness expansion. *Phys. Rev. Lett.* **109**, 160404 (2012).
61. P. Suppes, M. Zanotti, When are probabilistic explanations possible? *Synthese* **48**, 191–199 (1981).
62. A. Garg, N. D. Mermin, Farkas's Lemma and the nature of reality: Statistical implications of quantum correlations. *Found. Phys.* **14**, 1–39 (1984).
63. M. Zukowski, C. Brukner, Bell's theorem for general N-qubit states. *Phys. Rev. Lett.* **88**, 210401 (2002).
64. A. A. Klyachko, M. A. Can, S. Binicioğlu, A. S. Shumovsky, Simple test for hidden variables in spin-1 systems. *Phys. Rev. Lett.* **101**, 020403 (2008).
65. J. Barrett *et al.*, Nonlocal correlations as an information-theoretic resource. *Phys. Rev. A* **71**, 022101 (2005).
66. J. Barrett, S. Pironio, Popescu-Rohrlich correlations as a unit of nonlocality. *Phys. Rev. Lett.* **95**, 140401 (2005).
67. N. S. Jones, L. Masanes, Interconversion of nonlocal correlations. *Phys. Rev. A* **72**, 052312 (2005).
68. J. V. Kujala, E. N. Dzhafarov, J. A. Larsson, Necessary and sufficient conditions for an extended noncontextuality in a broad class of quantum mechanical systems. *Phys. Rev. Lett.* **115**, 150401 (2015).
69. E. N. Dzhafarov, J. V. Kujala, J. A. Larsson, Contextuality in three types of quantum-mechanical systems. *Found. Phys.* **45**, 762–782 (2015).
70. N. D. Mermin, Quantum mysteries revisited. *Am. J. Phys.* **58**, 731–734 (1990).
71. M. O. Renou *et al.*, Genuine quantum nonlocality in the triangle network. *Phys. Rev. Lett.* **123**, 140401 (2019).
72. N. Gisin, *Quantum Chance: Nonlocality, Teleportation and Other Quantum Marvels* (Springer, 2014).
73. S. Popescu, D. Rohrlich, Quantum nonlocality as an axiom. *Found. Phys.* **24**, 379 (1994).

PNAS
www.pnas.org

Supplementary Information for

Violations of locality and free choice are equivalent resources in Bell experiments

Pawel Blasiak, Emmanuel M. Pothos, James M. Yearsley, Christoph Gallus and Ewa Borsuk

Corresponding Author: Pawel Blasiak.
E-mail: pawel.blasiak@ifj.edu.pl

This PDF file includes:

Supporting Information Text

In this Supplement we prove two technical results **Lemma 2** and **Lemma 3** from the **Methods** section in the main manuscript. We also give an alternative self-standing proof of **Theorem 3**.

[The numbering of equations follows the main text.]

PROOF OF LEMMA 2

For reference we quote **Lemma 2** from the main manuscript that we will shortly prove. Here, the choice of settings is arbitrary $x, y \in \mathfrak{M} = \{1, 2, \ldots, M\}$.

Lemma 2. *For any behaviour $\{P_{ab|xy}\}_{xy}$ and distribution of settings P_{xy} there exists a local hidden variable model (LHV) which fully violates the freedom of choice assumption (i.e., if $\tilde{\Lambda}$ is the relevant HV space, then we have $\tilde{\Lambda} = \tilde{\Lambda}_L = \tilde{\Lambda}_{NF}$; cf. Eqs. (10) and (11)).*

Proof. Let us first introduce an auxiliary pair of indices $uv \in \mathfrak{M} \times \mathfrak{M}$ where $\mathfrak{M} = \{1, 2, \ldots, M\}$ and for each pair uv define a trivial behaviour $\{\tilde{P}^{(uv)}_{ab|xy}\}_{xy}$ which consists of $M \times M$ copies of the same distribution

$$\tilde{P}^{(uv)}_{ab|xy} \;:=\; P_{ab|uv} \quad \text{for all } xy \in \mathfrak{M} \times \mathfrak{M}\,, \tag{34}$$

i.e., this holds independently of the choice of settings xy. Then, for each such behaviour there exists an LHV model. To see this, for a given pair of indices uv it is enough to consider a four-element HV space

$$\tilde{\Lambda}^{(uv)} \;:=\; \{(\alpha, \beta) : \alpha, \beta = \pm 1\}\,, \tag{35}$$

with probability distributions defined in the following way

$$\tilde{P}^{(uv)}_{\lambda} \;:=\; \tilde{P}_{\alpha\beta|uv} \quad \text{and} \quad \tilde{P}^{(uv)}_{ab|xy\lambda} \;:=\; \delta_{a=\alpha} \cdot \delta_{b=\beta}\,, \tag{36}$$

where $\lambda = (\alpha, \beta) \in \tilde{\Lambda}^{(uv)}$. By construction, the locality condition Eq. (8) is satisfied for all λ, i.e., $\tilde{\Lambda}^{(uv)} = \tilde{\Lambda}^{(uv)}_L$, and the marginals reproduce the behaviour in Eq. (34)

$$\tilde{P}^{(uv)}_{ab|xy} \;=\; \sum_{\lambda \in \tilde{\Lambda}^{(uv)}} \tilde{P}^{(uv)}_{ab|xy\lambda} \cdot \tilde{P}^{(uv)}_{\lambda}\,. \tag{37}$$

Those LVH models of auxiliary behaviours $\{\tilde{P}^{(uv)}_{ab|xy}\}_{xy}$ can be used to construct another local model on a compound HV space

$$\tilde{\Lambda} \;:=\; \biguplus_{u,v \,\in\, \mathfrak{M}} \tilde{\Lambda}^{(uv)}\,, \tag{38}$$

with conditional probabilities on each component left unchanged, i.e., for each $u, v \in \mathfrak{M}$ we have

$$\tilde{P}_{ab|xy\lambda} \;:=\; \tilde{P}^{(uv)}_{ab|xy\lambda} \qquad \text{for } \lambda \in \tilde{\Lambda}^{(uv)}\,. \tag{39}$$

Note that in this way the *locality* condition is retained, i.e., the factorisation of Eq. (8) holds for every $\lambda \in \tilde{\Lambda}$ (cf. Eq. (36)), and we have $\tilde{\Lambda}_L = \biguplus_{u,v \,\in\, \mathfrak{M}} \tilde{\Lambda}^{(uv)}_L = \tilde{\Lambda}$. Then we impose the conditional distributions $\tilde{P}_{xy|\lambda}$ in the form

$$\tilde{P}_{xy|\lambda} \;:=\; \delta_{x=u} \cdot \delta_{y=v} \qquad \text{for } \lambda \in \tilde{\Lambda}^{(uv)}\,. \tag{40}$$

Clearly, this violates the *free choice* assumption Eq. (9) for every $\lambda \in \tilde{\Lambda}$ (since the knowledge of λ fully determines the settings xy), i.e., we have $\tilde{\Lambda}_{NF} = \tilde{\Lambda}$. Finally, we define the distribution of $\tilde{P}_{\lambda}$ on the whole HV space $\tilde{\Lambda}$ by the following condition

$$\tilde{P}_{\lambda} \;:=\; \tilde{P}^{(uv)}_{\lambda} \cdot P_{uv} \qquad \text{for } \lambda \in \tilde{\Lambda}^{(uv)}\,, \tag{41}$$

where P_{uv} is the desired distribution of settings (which for sake of generality is left unspecified). Observe that Eqs. (40) and (41) allow us to reverse the conditioning, i.e., Bayes' rule gives

$$\tilde{P}_{\lambda|xy} \;:=\; \begin{cases} \tilde{P}^{(xy)}_{\lambda} & \text{if } \lambda \in \tilde{\Lambda}^{(xy)}\,, \\ 0 & \text{otherwise}\,. \end{cases} \tag{42}$$

Having defined in Eqs. (38)-(41) all components of the HV model (*local* and fully violating *freedom of choice*) we need to check whether it correctly reconstructs some given behaviour $\{P_{ab|xy}\}_{xy}$ and distribution of settings P_{xy} via Eqs. (6) and (7). The following calculation

$$\sum_{\lambda \in \tilde{\Lambda}} \tilde{P}_{ab|xy\lambda} \cdot \tilde{P}_{\lambda|xy} \overset{(42)}{=} \sum_{\lambda \in \tilde{\Lambda}^{(xy)}} \tilde{P}_{ab|xy\lambda} \cdot \tilde{P}^{(xy)}_{\lambda} \tag{43}$$

$$\overset{(39)}{=} \sum_{\lambda \in \tilde{\Lambda}^{(xy)}} \tilde{P}^{(xy)}_{ab|xy\lambda} \cdot \tilde{P}^{(xy)}_{\lambda} \tag{44}$$

$$\overset{(37)}{=} \tilde{P}^{(xy)}_{ab|xy} \overset{(34)}{=} P_{ab|xy}\,, \tag{45}$$

confirms the validity of Eq. (6). Similarly, we check that

$$\sum_{\lambda \in \tilde{\Lambda}} \tilde{P}_{xy|\lambda} \cdot \tilde{P}_{\lambda} \overset{(40)}{=} \sum_{\lambda \in \tilde{\Lambda}^{(xy)}} \tilde{P}_{\lambda} \tag{46}$$

$$\overset{(41)}{=} \sum_{\lambda \in \tilde{\Lambda}^{(xy)}} \tilde{P}^{(xy)}_{\lambda} \cdot P_{xy} \;=\; P_{xy}\,, \tag{47}$$

in compliance with Eq. (7). This concludes the proof. □

PROOF OF LEMMA 3

In the following we consider a Bell experiment with binary settings $x, y \in \mathfrak{M} = \{0, 1\}$.

Preliminaries. We begin with a useful observation about the CHSH expressions in Eqs. (2)-(5).

Remark 3. *We have $|S_i| + |S_j| \leqslant 4$ whenever $i \neq j$.*

Proof. This is a direct consequence of a simple algebraic inequality involving the four expressions $S_1, \ldots, S_4$ which follow the pattern of Eqs. (2)-(5), i.e., $S_1 = a + b + c - d$, $S_2 = a + b - c + d$, $S_3 = a - b + c + d$ and $S_4 = -a + b + c + d$. It is straightforward to check that for $-1 \leqslant a, b, c, d \leqslant 1$, whenever $i \neq j$ we have $\pm S_i \pm S_j \leqslant 4$ and $\pm S_i \mp S_j \leqslant 4$ (since in each case four out of eight terms cancel out). Clearly, in compact form this is equivalent to the inequality $|S_i| + |S_j| \leqslant 4$ which concludes the proof. □

As an immediate consequence of **Remark 3**, note that *for any behaviour $\{P_{ab|xy}\}_{xy}$ at most one out of four CHSH inequalities in Eq. (1) can be violated at the same time.*

Without loss of generality, in the following proof we consider a given behaviour $\{P_{ab|xy}\}_{xy}$ with binary choice of settings $x, y \in \mathfrak{M} = \{0, 1\}$ for which only the first CHSH inequality is violated, i.e., we assume that

$$S_{max} \;=\; S_1 \;>\; 2\,, \tag{48}$$

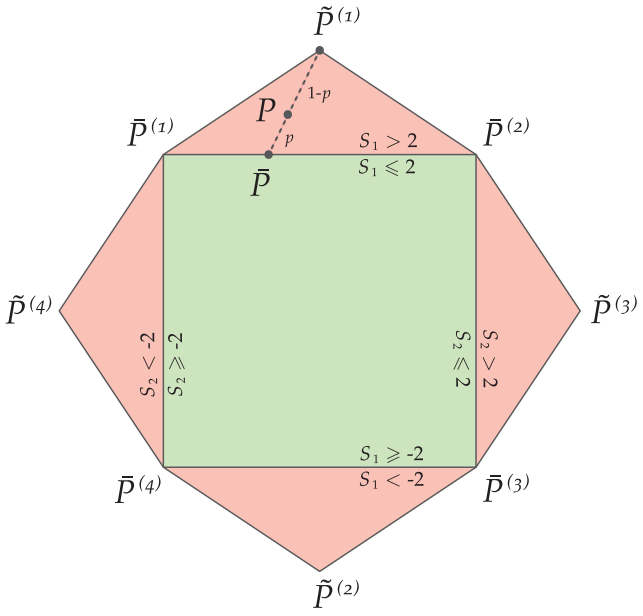


Fig. S1. Decomposition in Lemma 3. A two-dimensional section of the *non-signalling* polytope for the two-setting and two-outcome Bell scenario. It is spanned by sets of *local* deterministic behaviours $\bar{P}^{(j)}$ and *non-local* PR-boxes $\tilde{P}^{(k)}$. The subset of *local behaviours* (depicted as the inner green rectangle) is fully characterised by the CHSH inequalities Eq. (1). For each non-signalling behaviour P which breaks the local bound, one finds a convex decomposition $P = p \cdot \bar{P} + (1-p) \cdot \tilde{P}^{(k)}$ with $p = \frac{1}{2}(4 - S_{max})$ for some local behaviour $\bar{P}$ and a PR-box $\tilde{P}^{(k)}$. On the picture, we have $S_{max} = S_1 > 2$ and $k = 1$.

and consequently $|S_i| < 2$ for $i = 2,3,4$. It is straightforward to appreciate that all other cases reduce to this one by relabelling the indices.*

Decomposition of Lemma 3 (proof). Let us start by recalling the concept of a PR-box (due to Popescu-Rohrlich [1]) which is a non-signalling behaviour $\{\tilde{P}_{ab|xy}\}_{xy}$ saturating the CHSH expressions. To be more precise, only one of the expressions Eqs. (2)-(5) can reach its maximal/minimal value $\tilde{S}_i = \pm 4$ and then the rest must be equal to zero (cf. **Remark 3**). There are eight possible PR-boxes out of which we select just two cases for illustration

$$\tilde{P}^{(1)}_{ab|xy} = \begin{cases} 1/2 & \text{for } a \oplus b = x \cdot y\,, \\ 0 & \text{otherwise}\,, \end{cases} \qquad [49]$$

and

$$\tilde{P}^{(2)}_{ab|xy} = \begin{cases} 0 & \text{for } a \oplus b = x \cdot y\,, \\ 1/2 & \text{otherwise}\,. \end{cases} \qquad [50]$$

All the remaining boxes $\{\tilde{P}^{(k)}_{ab|xy}\}_{xy}$ for $k = 3,\ldots,8$ obtain by relabelling the indices x, y. These two behaviours saturate the first CHSH expression Eq. (2), i.e.

$$\begin{aligned} \tilde{S}^{(1)}_1 &= +4\,, \quad & \tilde{S}^{(1)}_i &= 0 \quad \text{for } i = 2,3,4\,, \\ \tilde{S}^{(2)}_1 &= -4\,, \quad & \tilde{S}^{(2)}_i &= 0 \quad \text{for } i = 2,3,4\,, \end{aligned} \qquad [51]$$

and similarly the remaining six boxes saturate the other three CHSH expressions Eqs. (3)-(5) (in particular we have $\tilde{S}^{(k)}_1 = 0$ for $k = 3,\ldots,8$). In what follows, PR-boxes will be used because of their role in the description of the non-signalling polytope in a two-setting and two-outcome Bell scenario [2].

For reference let us quote **Lemma 3** from the main manuscript that we will shortly prove; see [3] for a related result. See Fig. S1 for illustration.

Lemma 3. *Each non-signalling behaviour $\{P_{ab|xy}\}_{xy}$ with binary settings $x, y \in \mathfrak{M} = \{0,1\}$ can be decomposed as a convex mixture of a local behaviour $\{\bar{P}_{ab|xy}\}_{xy}$ and a PR-box $\{\tilde{P}_{ab|xy}\}_{xy}$ in the form*

$$P_{ab|xy} = p \cdot \bar{P}_{ab|xy} + (1-p) \cdot \tilde{P}_{ab|xy}\,, \qquad [52]$$

with $p = \frac{1}{2}(4 - S_{max})$ for all $x, y \in \{0,1\}$.
[In the case $S_1 > 2$, the PR-box is that of Eq. (49)].

Here, local behaviour means existence of an LHV+FHV model of $\{\bar{P}_{ab|xy}\}_{xy}$ and $S_{max} = \max\{|S_i| : i = 1,\ldots,4\}$.

Proof. Let us consider a given non-signalling behaviour $\{P_{ab|xy}\}_{xy}$ such that $S_1 > 2$, and define the following four distributions

$$\bar{P}_{ab|xy} := \frac{1}{p}\, P_{ab|xy} - \frac{1-p}{p}\, \tilde{P}^{(1)}_{ab|xy}\,, \qquad [53]$$

for $x, y \in \{0,1\}$, where $\{\tilde{P}^{(1)}_{ab|xy}\}_{xy}$ is the PR-box defined in Eq. (49). In order to prove **Lemma 3** it is enough to show that

(a) $\bar{P}_{ab|xy}$ are well-defined probability distributions for each $x, y \in \{0,1\}$, and

(b) $\{\bar{P}_{ab|xy}\}_{xy}$ defines a local behaviour,

for the choice $p = \frac{1}{2}(4 - S_1)$.

Ad. (a). The normalisation condition $\sum_{ab} \bar{P}_{ab|xy} = 1$ is trivially satisfied, since $P_{ab|xy}$ and $\tilde{P}^{(1)}_{ab|xy}$ are normalised. It only remains to check that $\bar{P}_{ab|xy} \geqslant 0$.

It is known that the non-signalling polytope for the two-setting and two-outcome scenario is characterised by $24 = 16 + 8$ extremal points [2] (cf. Fig. S1):

$\{\bar{P}^{(j)}_{ab|xy}\}_{xy}$ – sixteen local deterministic behaviours for $j = 1,\ldots,16$,

$\{\tilde{P}^{(k)}_{ab|xy}\}_{xy}$ – eight PR-boxes for $k = 1,\ldots,8$.

This means that we can decompose $\{P_{ab|xy}\}_{xy}$ as a convex combination

$$P_{ab|xy} = \sum_{j=1}^{16} p_j \cdot \bar{P}^{(j)}_{ab|xy} + \sum_{k=1}^{8} q_k \cdot \tilde{P}^{(k)}_{ab|xy} \qquad [54]$$

with proper normalisation $\sum_{j=1}^{16} p_j + \sum_{k=1}^{8} q_k = 1$ and $p_j, q_k \geqslant 0$ for all j, k. We can use Eq. (54) to calculate the first CHSH expression Eq. (2), which by linearity gives

$$S_1 = \sum_{j=1}^{16} p_j \cdot \bar{S}^{(j)}_1 + \sum_{k=1}^{8} q_k \cdot \tilde{S}^{(k)}_1\,, \qquad [55]$$

where $\bar{S}^{(j)}_1$ and $\tilde{S}^{(k)}_1$ correspond to $\{\bar{P}^{(j)}_{ab|xy}\}_{xy}$ and $\{\tilde{P}^{(k)}_{ab|xy}\}_{xy}$ respectively. For this particular choice of distributions, we have $|\bar{S}^{(j)}_1| \leqslant 2$ (for local behaviours) and Eq. (51) (for PR-boxes), from which the following bound is obtained

$$\begin{aligned} S_1 &= \sum_{j=1}^{16} p_j \cdot \bar{S}^{(j)}_1 + 4 \cdot q_1 - 4 \cdot q_2 \\ &\leqslant 2 \cdot \sum_{j=1}^{16} p_j + 4 \cdot q_1 \leqslant 2 + 2 \cdot q_1\,. \end{aligned} \qquad [56]$$

*To change signs: $(a \to -a) \Rightarrow (S_i \to -S_i)$. To switch between expressions: $(x \to 1-x) \Rightarrow (S_1 \leftrightarrow S_3)$, $(y \to 1-y) \Rightarrow (S_1 \leftrightarrow S_2)$ and $(x \to 1-x \;\&\; y \to 1-y) \Rightarrow (S_1 \leftrightarrow S_4)$.

Note that in the last inequality we used the normalisation $\sum_{j=1}^{16} p_j + \sum_{k=1}^{8} q_k = 1$. Thus, in the decomposition of Eq. (54) the coefficient $q_1 \geqslant \frac{1}{2}S_1 - 1$. A quick comparison of Eqs. (53) and (54) reveals that such a defined distribution is well-defined, i.e., $\bar{P}_{ab|xy} \geqslant 0$, as long as $q_1 \geqslant 1 - p$. This is true for $p \geqslant \frac{1}{2}(4 - S_1)$.

Ad. (b). Since both sets of behaviours $\{P_{ab|xy}\}_{xy}$ and the PR-box $\{\tilde{P}^{(1)}_{ab|xy}\}_{xy}$ are non-signalling, then $\{\bar{P}_{ab|xy}\}_{xy}$ defined in Eq. (53) is non-signalling too. Therefore, by virtue of Fine's theorem [4, 5], it suffices to check that all CHSH inequalities are satisfied $|\bar{S}_i| \leqslant 2$.

By linearity of the CHSH expresions, from Eq. (53) we get $\bar{S}_i = \frac{1}{p} S_i - \frac{1-p}{p} \tilde{S}^{(1)}_i$. It follows that

$$|\bar{S}_i| = \frac{|S_i - (1-p)\,\tilde{S}^{(1)}_i|}{p} \overset{(51)}{=} \begin{cases} \frac{|S_1 - 4\,(1-p)|}{p} & \text{for } i = 1\,, \\ \frac{|S_i|}{p} \leqslant \frac{4-|S_1|}{p} & \text{for } i \neq 1\,, \end{cases} \qquad [57]$$

where the last inequality for $i \neq 1$ is due to **Remark 3**. Now, it is straightforward to check that for $p = \frac{1}{2}(4 - S_1)$ all four CHSH expressions satisfy the local bound, i.e. we have $|\bar{S}_1| = 2$ and $|\bar{S}_i| \leqslant 2$ for $i \neq 1$.

□

DIRECT PROOF OF THEOREM 3

We start by considering a wide class of upper bounds on the measure free choice μ_F. This result will be used to prove **Theorem 3** with the help of correlations saturating the so called chained Bell inequalities (for a special choice of measurement settings on a Bell state).

Upper bounds on measure of free choice μ_F. Let $x, y \in \mathfrak{M}$ with the choice of settings $\mathfrak{M}$ unspecified for the time being. Our analysis will proceed with algebraic expressions which take the following form

$$S = \sum_{x,y} \alpha_{xy} \cdot \langle ab \rangle_{xy}\,, \qquad [58]$$

where $\alpha_{xy} \in \{0, \pm 1\}$ are some arbitrary coefficients and $\langle ab \rangle_{xy} \equiv \sum_{a,b} ab\, P_{ab|xy}$ is the usual correlation function. For each such expression S we will look at two auxiliary quantities. The first one counts the number of non-zero terms in Eq. (58), i.e.,

$$S^{\#} = \sum_{x,y} |\alpha_{xy}|\,. \qquad [59]$$

The second one is the upper bound of Eq. (58) with the correlation coefficients replaced by a product, i.e.,[†]

$$S^{loc} = \max_{\substack{-1 \leqslant a_x \leqslant 1 \\ -1 \leqslant b_y \leqslant 1}} \sum_{x,y} \alpha_{xy} \cdot a_x b_y\,. \qquad [60]$$

Note that both $S^{\#}$ and S^{loc} depend only on the coefficients α_{xy} (and not on the correlation functions $\langle ab \rangle_{xy}$).

In the following, we consider an arbitrary HV model of a given set behaviour $\{P_{ab|xy}\}_{xy}$ obtained in a Bell experiment (with $x, y \in \mathfrak{M}$). The decomposition Eq. (11) entails splitting of Eq. (6) into two components

$$P_{ab|xy} = \sum_{\lambda \in \Lambda_F} P_{ab|xy\lambda} \cdot P_\lambda + \sum_{\lambda \in \Lambda_{NF}} P_{ab|xy\lambda} \cdot P_{\lambda|xy}\,, \qquad [61]$$

which describe fundamentally different situations *with* and *without* freedom of choice, i.e. corresponding to the hidden variable λ being respectively in the *free domain* (Λ_F) or *non-free domain* (Λ_{NF}). This can be used to write the correlation coefficients in the form

$$\langle ab \rangle_{xy} = \sum_{\lambda \in \Lambda_F} \langle ab \rangle_{xy\lambda} \cdot P_\lambda + \sum_{\lambda \in \Lambda_{NF}} \langle ab \rangle_{xy\lambda} \cdot P_{\lambda|xy}\,, \qquad [62]$$

where $\langle ab \rangle_{xy\lambda} \equiv \sum_{a,b} ab\, P_{ab|xy\lambda}$ is the correlation coefficient for fixed λ. Hence, the expression S defined in Eq. (58) splits into two parts

$$S = \sum_{\lambda \in \Lambda_F} \Big(\sum_{x,y} \alpha_{xy} \cdot \langle ab \rangle_{xy\lambda} \Big) \cdot P_\lambda + \qquad [63]$$

$$\sum_{\lambda \in \Lambda_{NF}} \Big(\sum_{x,y} \alpha_{xy} \cdot \langle ab \rangle_{xy\lambda} \cdot P_{\lambda|xy} \Big)\,. \qquad [64]$$

Note that in the first term probability P_λ factors out due to the condition in Eq. (11) defining the *free domain* Λ_F. So far, the causal structure of the HV model has not been constrained.

Now, we make the *locality assumption* Eq. (8) which entails the following factorisation $\langle ab \rangle_{xy\lambda} = \langle a \rangle_{x\lambda} \langle b \rangle_{y\lambda}$, with $\langle a \rangle_{x\lambda} = \sum_a a\, P_{a|x\lambda}$ and $\langle b \rangle_{y\lambda} = \sum_b b\, P_{b|y\lambda}$. Then the expression in Eq. (63) is bounded by S^{loc} of Eq. (60), i.e.,

$$\text{Eq. (63)} \leqslant \sum_{\lambda \in \Lambda_F} S^{loc} \cdot P_\lambda = S^{loc} \cdot \sum_{\lambda \in \Lambda_F} P_\lambda\,. \qquad [65]$$

For the expression in Eq. (64) we use $S^{\#}$ of Eq. (59) to make the following estimate

$$\text{Eq. (64)} \leqslant \sum_{x,y} |\alpha_{xy}| \sum_{\lambda \in \Lambda_{NF}} P_{\lambda|xy} \overset{(67)}{=} \sum_{x,y} |\alpha_{xy}| \cdot \Big(1 - \sum_{\lambda \in \Lambda_F} P_\lambda \Big)$$
$$\leqslant S^{\#} \cdot \Big(1 - \sum_{\lambda \in \Lambda_F} P_\lambda \Big)\,, \qquad [66]$$

where we use the fact that conditional probabilities $P_{\lambda|xy}$ are normalised and $\Lambda = \Lambda_F \cup \Lambda_{NF}$ (cf. Eq. (11)), which entails

$$1 = \sum_{\lambda \in \Lambda} P_{\lambda|xy} = \sum_{\lambda \in \Lambda_F} P_\lambda + \sum_{\lambda \in \Lambda_{NF}} P_{\lambda|xy} \qquad [67]$$

for each $x, y \in \mathfrak{M}$. Putting everything together, from Eqs. (63)/(64) and (65)/(66) we get the inequality

$$S \leqslant S^{loc} \cdot \sum_{\lambda \in \Lambda_F} P_\lambda + S^{\#} \cdot (1 - \sum_{\lambda \in \Lambda_F} P_\lambda)\,,$$

which provides the following bound on the amount of free choice in a given LHV model (see **Remark 1**)

$$\sum_{\lambda \in \Lambda_F} P_\lambda \leqslant \frac{S^{\#} - S}{S^{\#} - S^{loc}}\,. \qquad [68]$$

Since this inequality should be satisfied by any LHV model reproducing given behaviour $\{P_{ab|xy}\}_{xy}$ in a Bell experiment (irrespective of the distribution of settings P_{xy}), then we get upper bound on the the measure of free choice defined in Eq. (13)

$$\mu_F \leqslant \frac{S^{\#} - S}{S^{\#} - S^{loc}}\,. \qquad [69]$$

[†] A closer inspection of Eq. (60) reveals the maximum is attained at one of the extremal points $a_x, b_y = \pm 1$ (since the expression $\sum_{x,y} \alpha_{xy} \cdot a_x b_y$ is a linear function for each variable a_x, b_y).

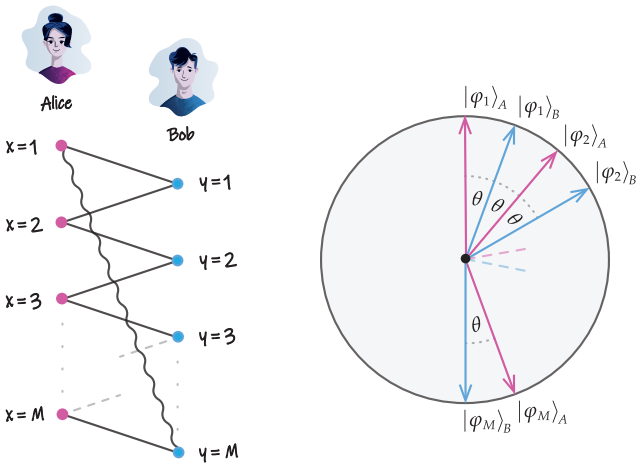


Fig. S2. Chained Bell inequalities. On the left, there is a schematic illustration of the chained expression in Eq. (70) with each one of $2M$ terms $\langle ab\rangle_{xy}$ represented by the line $x-y$. Solid lines correspond to the '$+$' signs and the wavy line corresponds to the '$-$' sign. On the right, the $\hat{x}-\hat{z}$ section of the Bloch sphere concerns measurement directions defined in Eqs. (72) and (73). Alice measures along the directions $|\varphi_x\rangle_A$ (depicted in magenta) and Bob measures along the directions $|\varphi_y\rangle_B$ (depicted in blue), with $x,y=1,2,\ldots,M$ and $\theta=\frac{\pi}{2M-1}$.

Crucially, the bound in Eq. (69) holds for any expression S as defined in Eq. (58). Let us emphasise that the numbers $S^{\#}$ and S^{loc} are determined solely by the choice of coefficients α_{xy}, while S itself depends on the observed experimental statistics $\{P_{ab|xy}\}_{xy}$ through the correlation coefficients $\langle ab\rangle_{xy}$, see Eqs. (58) - (60).

Proof of Theorem 3. Quantum correlations are stronger than the classical ones, but much weaker than what one might possibly imagine [1, 6]. For the proof of **Theorem 3** is however enough to exclude all freedom of choice for any local account of a maximally entangled state when the number of settings $x,y\in\mathfrak{M}=\{1,2,\ldots,M\}$ goes to infinity $M\to\infty$.

Let us take expression of Eq. (58) in a special chained form [7, 8]

$$\begin{aligned} S_M &= \langle ab\rangle_{11}+\langle ab\rangle_{21}+\langle ab\rangle_{22}+\langle ab\rangle_{32}+\ldots \\ &\quad \ldots+\langle ab\rangle_{MM}-\langle ab\rangle_{1M}\,, \end{aligned} \qquad [70]$$

which follows the pattern illustrated in Fig. S2 (on the left). From simple term counting we have $S_M^{\#}=2M$. Now, if locality is imposed this leads to the so called *chained Bell inequalities*: $S_M\leqslant 2(M-1)$, see [7, 8]. It is straightforward to see that Eq. (60) saturates to $S_M^{loc}=2(M-1)$.[‡] This means that for the expression Eq. (70) the upper bound on the amount of free choice in Eq. (69) becomes

$$\mu_F \leqslant \frac{2M-S_M}{2M-2(M-1)} = M-\tfrac{1}{2}S_M\,. \qquad [71]$$

Now, we will assume that the correlations in a Bell experiment are described by the quantum statistics given by the Bell state $|\Phi_+\rangle\equiv\frac{1}{\sqrt{2}}(|0\rangle_A|0\rangle_B+|1\rangle_A|1\rangle_B)$ and the choice of settings x,y implements projective measurements in the respective bases $\{|\varphi_x\rangle,|\varphi_x\rangle^\perp\}$ for Alice and $\{|\varphi_y\rangle,|\varphi_y\rangle^\perp\}$ for Bob, where

$$|\varphi_x\rangle_A = \cos(x-1)\theta\,|0\rangle_A+\sin(x-1)\theta\,|1\rangle_A\,, \qquad [72]$$

$$|\varphi_y\rangle_B = \cos(y-\tfrac{1}{2})\theta\,|0\rangle_B+\sin(y-\tfrac{1}{2})\theta\,|1\rangle_B\,, \qquad [73]$$

for $x,y\in\mathfrak{M}\equiv\{1,2,\ldots,M\}$ and $\theta=\frac{\pi}{2M-1}$. See Fig. S2 (on the right) for illustration. Then all terms in Eq. (70) become equal to $\langle ab\rangle_{xy}=\cos\theta$ except for the last one which is equal to $\langle ab\rangle_{1M}=-1$, and hence we get

$$\begin{aligned} S_M &= (2M-1)\cos\theta+1 \\ &\gtrsim (2M-1)\Big(1-\tfrac{\pi^2}{2(2M-1)^2}\Big)+1\,. \end{aligned} \qquad [74]$$

Substituted into Eq. (71), this gives

$$\mu_F \leqslant \frac{\pi^2}{4(2M-1)} \xrightarrow[M\to\infty]{} 0\,. \qquad [75]$$

This justifies our claim about freedom of choice reduced to zero in any local account of the quantum state $|\Phi_+\rangle$. Generalisation to any maximally entangled two-qubit state is straightforward.

References

1. S Popescu, D Rohrlich, Quantum Nonlocality as an Axiom. *Found. Phys.* **24**, 379 (1994).
2. V Scarani, *Bell Nonlocality*. (Oxford University Press), (2019).
3. S Pironio, Violations of Bell inequalities as lower bounds on the communication cost of nonlocal correlations. *Phys. Rev. A* **68**, 062102 (2003).
4. A Fine, Hidden Variables, Joint Probability, and the Bell Inequalities. *Phys. Rev. Lett.* **48**, 291–295 (1982).
5. JJ Halliwell, Two proofs of Fine's theorem. *Phys. Lett. A* **378**, 2945–2950 (2014).
6. S Popescu, Nonlocality beyond quantum mechanics. *Nat. Phys.* **10**, 264 (2014).
7. PM Pearle, Hidden-Variable Example Based upon Data Rejection. *Phys. Rev. D* **2**, 1418 (1970).
8. SL Braunstein, CM Caves, Wringing out better Bell inequalities. *Ann. Phys.* **202**, 22 (1990).

[‡] Locality entails factorisation $\langle ab\rangle_{xy}=\langle a\rangle_x\langle b\rangle_y=:a_xb_y$. Then $S_M=(a_1+a_2)b_1+(a_2+a_3)b_2+\ldots+(a_M-a_1)b_M$. Suppose $a_x,b_y=\pm1$, then out of M terms taking values $0,\pm2$ at least one has to be equal to zero (if $a_M=a_1$ it is the last one, otherwise it is one of the remaining ones). In such a case $S_M\leqslant 2(M-1)$. This extends to intermediate values $-1\leqslant a_x,b_y\leqslant+1$, since S_M is a linear function for each of the variables and hence the global maximum is attained at the extremal points $a_x,b_y=\pm1$. Clearly, taking $a_x=b_y=1$ we get the maximum, i.e. $S_M^{loc}=2(M-1)$.

◇ ◇ ◇ ◇ ◇