Journal of Physics A:
Mathematical and
Theoretical

**PAPER**

# Consistency and causality of interconnected nonsignaling resources

To cite this article: Peter Bierhorst 2024 *J. Phys. A: Math. Theor.* **57** 425301

View the article online for updates and enhancements.

You may also like

- Bell inequalities tailored to the Greenberger–Horne–Zeilinger states of arbitrary local dimension
R Augusiak, A Salavrakos, J Tura et al.

- Translationally invariant multipartite Bell inequalities involving only two-body correlators
J Tura, A B Sainz, T Vértesi et al.

- Measurement dependent locality
Gilles Pütz and Nicolas Gisin

This content was downloaded from IP address 98.163.251.79 on 27/06/2025 at 18:33

IOP Publishing Journal of Physics A: Mathematical and Theoretical

J. Phys. A: Math. Theor. **57** (2024) 425301 (28pp) https://doi.org/10.1088/1751-8121/ad7c9f

# Consistency and causality of interconnected nonsignaling resources

**Peter Bierhorst**

Mathematics Department, University of New Orleans, 2000 Lakeshore Drive, New Orleans, LA 70148, United States of America

E-mail: plbierho@uno.edu

Received 13 June 2024; revised 26 August 2024
Accepted for publication 18 September 2024
Published 30 September 2024

**Abstract**

This paper examines networks of $n$ measuring parties sharing $m$ nonsignaling resources that can be locally wired together: that is, each party follows a scheme to measure the resources in a cascaded fashion with inputs to later resources possibly depending on outputs of earlier-measured ones. A specific framework is provided for studying probability distributions arising in such networks, and this framework is used to directly prove some accepted, but often only implicitly invoked, facts: there is a uniquely determined and well-defined joint probability distribution for the outputs of all resources shared by the parties, and this joint distribution is nonsignaling. It is furthermore shown that is often sufficient to restrict consideration to only extremal nonsignaling resources when considering features and properties of such networks. Finally, the framework illustrates how the physical theory of nonsignaling boxes and local wirings is *causal*, supporting the applicability of the inflation technique to constrain such models. For an application, we probe the example of $(3,2,2)$ inequalities that witness genuine three-party nonlocality according to the local-operations-shared-randomness definition, and show how all other examples can be derived from that of Mao *et al* (2022 *Phys. Rev. Lett.* **129** 150401).

Keywords: causality, wired nonsignaling resources, genuine multipartite nonlocality

## 1. Introduction: nonsignaling resources and networks

Quantum mechanics is *nonlocal* in the sense that certain quantum experiments involving spatially separated measuring parties do not admit a local hidden variable description [1, 2]. Quantum nonlocality experiments have been performed under strict conditions [3–6], confirming the phenomenon.

© 2024 IOP Publishing Ltd.
All rights, including for text and data mining, AI training, and similar technologies, are reserved. 

Quantum nonlocality experiments, while demonstrating strange correlations between distant parties, still satisfy a condition known as the *nonsignaling* condition: a measuring party cannot exploit a quantum experiment to send signals to a spatially separated party. This assures compliance with special relativity in experiments where separated measurements can be performed near-simultaneously at great distances such that any signals would need to be traveling faster than the speed of light. The study of the non-signaling condition is of interest in various contexts, such as abstract definitions of nonlocality not invoking quantum mechanics [7], attempts to derive quantum mechanics from physical principles [8], and certifying unpredictability of random numbers under minimal assumptions [9].

The nonsignaling condition can be stated formally as follows: consider an experiment of two spatially separated parties named Alice and Bob, each of whom performs a measurement using an apparatus; each apparatus has a choice of measurement setting labeled (respectively) $X$ and $Y$ and provides a measurement outcome labeled (respectively) $A$ and $B$. Then the joint probability distribution $\mathbb{P}(A,B|X,Y)$ of outcomes conditioned on settings is *nonsignaling* if each party's marginal outcome probabilities are independent of settings choices of the other. Mathematically this can be expressed as the equalities

$$\begin{aligned} \mathbb{P}(A=a|X=x,Y=y) &= \mathbb{P}(A=a|X=x,Y=y') \\ \mathbb{P}(B=b|X=x,Y=y) &= \mathbb{P}(B=b|X=x',Y=y) \end{aligned} \tag{1}$$

which hold for all values $a,x,y$ and $x',y'$, where a marginal probability such as $\mathbb{P}(A=a|X=x,Y=y)$ is obtained from the joint distribution by the summation $\sum_b \mathbb{P}(A=a,B=b|X=x,Y=y)$. All distributions $\mathbb{P}$ obtainable with quantum mechanics satisfy the nonsignaling condition (1), but the converse is not true: there are distributions satisfying (1) that cannot be observed through measurements of entangled quantum states, such as the paradigmatic example of the Popescu–Rohrlich 'PR box' distribution [8]. Equation (1) can be generalized to scenarios of $n>2$ separated measuring parties, whereby it is stipulated that the outcome distribution of any given subset of the $n$ parties is required to be independent of the settings of the other parties.

An interesting question is what sort of probability distributions can be observed in a three-party experiment for which each pair of parties share bipartite nonsignaling resources satisfying (1)—possibly multiple such resources, allowing local 'wirings' whereby each party can access their resources in cascaded fashion and condition inputs provided to later resources on observed outputs from earlier ones. Early results on this question can be found in section IIIC of [10] and [7]. Recently, the question is of renewed interest in light of arguments [11, 12] that only three-party probability distributions that *cannot* be replicated by such underlying networks of bipartite resources—possibly supplemented with global shared randomness—should be considered *genuinely multipartite nonlocal* (GMNL). This approach resolves an anomaly [12] in earlier definitions of the GMNL concept [13–15] in which parallel independent two-party nonlocality experiments can be counterintuitively classified as GMNL. The new revised notion of GMNL is named in [16] as LOSR-GMNL, with LOSR standing for local operations and shared randomness; quantum measurements of the three-way entangled GHZ state [17] can exhibit LOSR-GMNL [11, 12, 18] and recent experiments [19–21] provide some initial evidence of the phenomenon.

Motivated in part by the LOSR-GMNL definition, this paper studies the general question of how to systematically model $n$-party conditional distributions, or *behaviors*, of the form $\mathbb{P}(\mathbf{A}_1,\ldots,\mathbf{A}_n|\mathbf{X}_1,\ldots,\mathbf{X}_n)$ that are induced as follows: a network of $m$ nonsignaling resources, each shared by a subset of the parties, is measured by the parties in cascaded fashion after each party $i$ receives a setting $\mathbf{X}_i$; then, each party's final outcome $\mathbf{A}_i$ is a function of the observed outputs from the resources. This scenario can be referred to as *nonsignaling resources with local wirings*. The study of behaviors obtained this way essentially reduces to the study of the joint distribution of all the resource outputs in the underlying network: $\mathbb{P}(\vec{A}_1,....,\vec{A}_m|\mathbf{X}_1,...,\mathbf{X}_n)$ where $\vec{A}_j$ refers to all the outputs of the $j$th shared resource.

The central goal of this work is thus to provide a framework for direct study of the joint probability distributions $\mathbb{P}(\vec{A}_1,....,\vec{A}_m|\mathbf{X}_1,...,\mathbf{X}_n)$ so as to 1) directly prove here some commonly accepted properties of such distributions, and 2) provide a foundation for rigorous future results about them. Regarding point (1), it has been accepted, often tacitly, that there *is* a well-posed joint distribution $\mathbb{P}(\vec{A}_1,....,\vec{A}_m|\mathbf{X}_1,...,\mathbf{X}_n)$ which is itself nonsignaling and causal (causality here roughly corresponding to an intuitive notion that the marginal distribution of a subset of parties is determined only by the resources they measure and how they measure them; quantum mechanics is an example of a causal theory but other more exotic theories may be causal as well). But considering that for wired *signaling* resources, a consistent joint distribution is not in general possible (see figure 1 of [15] for an example), it is good to be clear about why this is true when the resources are nonsignaling. Accordingly, this work rigorously demonstrates that the joint distribution induced by a network of wired nonsignaling resources is well-defined and itself nonsignaling. Moreover, while previous works such as [22–24] are cited in [12, 16] to justify the *causality* of the paradigm, these previous works are somewhat abstract and do not always address the point directly. The framework of this paper provides a clear foundation for demonstrating the causality of the theory of wired nonsignaling resources. An important consequence of the causality of the paradigm is that it enables use of the powerful inflation technique [25], which applies to causal theories.

Indeed, while this work reinforces the fact that that constraints derived from the inflation technique [25] are valid in constraining behaviors in networks of wired nonsignaling resources, it will provide an important foundation for deriving constraints satisfied by *only* these behaviors but possibly violated by different causal theories—an important example being scenarios allowing for entangled measurements and/or generalizations thereof. Since the inflation technique applies to all causal theories, it cannot readily be used to address this separation. Thus the framework introduced here for direct study of *just* wired nonsignaling resources will be useful in resolving the question of when/whether behaviors that can be observed in generalized probabilistic theories *with* entangled measurements (or generalized analogues thereof) can *not* be observed in networks where these are prohibited (such as nonsignaling resources with local wirings). This corresponds to the question of whether there are behaviors in regions $\mathcal{R}_3/\mathcal{R}_5$ in the Venn diagram of figure 2 of [26]; conjecture 1 in section V-C of [16] is an argument that $\mathcal{R}_5$ is nonempty. The question is somewhat subtle as some behaviors that would seemingly require entangled measurements—such as the device-independent certification of entangled measurements protocol of [27]—can be counterintuitively simulated with wired nonsignaling boxes [26]. Study of this region will increase our understanding of entangled measurements; it is also motivated by a variant definition [11] of LOSR-GMNL in which entangled measurements and generalizations thereof are not allowed for the class of behaviors that are classified as bipartite-only nonlocal.

Note that since quantum resources are nonsignaling, any constraint proved in this context will apply to a practical scenario of networks of quantum-achievable nonsignaling resources measured in cascaded fashion—the 'quantum box' paradigm of the set $QB_2$ in [26], which is directly relevant to the proposed definition of genuine network nonlocality given in [28].

The paper starts by defining nonsignaling resources and networks thereof in section 2, where a method for determining the joint distribution is formalized and shown to be consistent. Section 3 derives properties of joint distribution: the nonsignaling property, the ability to restrict attention to extremal nonlocal nonsignaling resources in certain cases (an important technique that was used in, for example, [11]), and a discussion of the causality of the framework which supports the applicability of the inflation technique.

The paper concludes with a case study example: inequalities witnessing LOSR-GMNL in the three-party scenario. It is shown that all known three-party inequalities with two settings per party and two outcomes per setting—the simplest possible scenario witnessing LOSR-GMNL (see section SM 3 of [26])—can be derived from the inequality of Mao *et al* [19] (which was obtained with the inflation technique, and so the causality results of this paper reinforce the applicability of this inequality to the paradigm of wired nonsignaling resources). These derivable inequalities include the inequality of Chao and Reichardt [18] as formulated in [11] (which is notable as the Chao–Reichardt inequality had previously only been derived directly within the paradigm of wired nonsignaling resources; by deriving it here as a consequence of the inequality of Mao *et al* we show it holds of the more broad class of causal theories), and inequality (1) of Cao *et al* [20]. A second inequality of Cao *et al*, which has an extra setting for one of the parties, can also be derived from that of Mao *et al*; a natural open question is whether different inequalities can be discovered in this scenario.

## 2. Nonsignaling resources: definition, a framework for studying networks, and consistency of the joint distribution

We are interested in behaviors $\mathbb{P}(\mathbf{A}_1,\ldots,\mathbf{A}_n|\mathbf{X}_1,\ldots,\mathbf{X}_n)$ that can be induced by underlying networks of nonsignaling resources. We will notate the distributions of the underlying network resources with $R$, as in $R(ABC|XYZ)$, to distinguish these from the final global distribution $\mathbb{P}$. It is also helpful to refer to the variables of the resource occurring in $R(\cdots|\cdots)$ as *outputs* and *inputs*, to distinguish them from the variables of the overall behavior $\mathbb{P}(\mathbf{A}_1,\ldots,\mathbf{A}_n|\mathbf{X}_1,\ldots,\mathbf{X}_n)$, for which we call $\mathbf{X_i}$ the *setting* and $\mathbf{A}_i$ the *outcome* or *final outcome*. In the next subsection, we introduce a formal definition of nonsignaling for an $n$ party resource $R$ that generalizes (1), and derive some important consequences of the nonsignaling condition.

### 2.1. Properties of nonsignaling resources

For an $n$-party resource $R(\cdots|\cdots)$, the nonsignaling condition is as follows: for each $j$ in $\{1,\ldots,n\}$ and each pair of possible values $x_j$ and $x_j'$ that the input choice $X_j$ can assume, we have

$$\sum_{a_j} R\left(A_1=a_1,\ldots,A_j=a_j,\ldots,A_n=a_n|X_1=x_1,\ldots,\underbrace{X_j=x_j}_{\text{change}},\ldots,X_n=x_n\right)$$
$$=\sum_{a_j} R\left(A_1=a_1,\ldots,A_j=a_j,\ldots,A_n=a_n|X_1=x_1,\ldots,\overbrace{X_j=x_j'}^{},\ldots,X_n=x_n\right) \tag{2}$$

for each fixed choice of $x_i$ among $i\neq j$. In words, this means that the conditional distribution of the $A_i$ *excluding* $A_j$ is independent of party $j$'s input choice. This represents the idea that one party (the $j$th) cannot signal to the rest through their choice of input. For the rest of the paper, we will use a shorthand in expressions like (2) whereby $R(\vec{a}|\vec{x})=R(a_1,\ldots,a_n|x_1,\ldots,x_n)$ is shorthand for $R(A_1=a_1,\ldots,A_n=a_n|X_1=x_1,\ldots,X_n)$.

A few points are worth mentioning before moving on. First, use of the conditional distribution notation $R(\vec{a}|\vec{x})$ suggests the existence of a joint distribution of all random variables

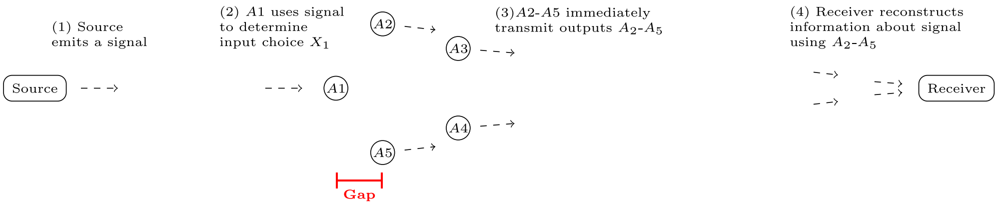


**Figure 1.** Faster-than-light signaling through violation of (2). In the scheme above, the signal distance between Source and Receiver is effectively shortened by the length of the span marked 'Gap,' with a small $\varepsilon$ correction owing to slight diagonality in the paths. The correction vanishes asymptotically as the distance to Receiver increases. If all dashed-line signals travel at the speed of light, signal information traverses the Source-Receiver span faster than the speed of light. Sources and receivers positioned along different diagonal axes motivate (2) for other choices of the signaling input $X_j$ ($j=1$ above). A similar figure with a regular $n$-gon motivates (2) for a set of $n$ parties.

comprising $\vec{A}$ and $\vec{X}$ from which the conditional probabilities are derived. However, in studies of networked nonsignaling resources, a full probability distribution of the inputs $\vec{X}$ is somewhat besides the point (even if it may exist)—we want to think of the inputs more as choices one can supply to the resources to which they respond. In this sense $R(\cdots|\cdots)$ is perhaps better thought of as a family of (unconditional) probability distributions of random variables $\vec{A}$, merely indexed by $\vec{x}$. Thus we avoid tacitly appealing to input probability distributions in the derivations below, so that consequences of (2) below could just as easily apply to $\vec{x}_n$-indexed families of (unconditional) probability distributions '$R_{\vec{x}_n}(\vec{A}_n)$' satisfying a suitably re-notated (2), while keeping the standard convention of conditional probability notation $R(\cdots|\cdots)$.

It also merits briefly discussing the physical motivation of (2). The justification via special relativity can be seen by considering a scenario where the $n$ parties are arranged at the vertices of a regular polygon, in which case any violation of (2) could result in a signal-speed boost in a particular direction: see figure 1 for an illustration in the case of five parties. The figure may make (2) seem incomplete, as there are of course other signalings that could be well motivated, such as many-to-one (running the figure in reverse), or a group of two adjacent parties signaling to the remaining three; conversely, certain other subset-to-subset signaling prohibitions might not be so clear how to intuitively justify based solely on special relativity considerations in the context of figure 1. However, it is known (see [10] section IIIA) that the other subset-to-subset signaling prohibitions can be derived mathematically from (2), and so once the (2) condition is accepted, one does not require new physical motivation to accept other nonsignaling conditions.

We now derive important consequences of (2). First, a more general prohibition on subset-to-subset signaling among the parties can be formulated in the following manner: For $1 \leqslant p \leqslant n$, let $\vec{a}_p$ denote $a_1,\ldots,a_p$ and $\vec{a}_q$ denote $a_{p+1},\ldots,a_n$, and let $\vec{x}_p$ and $\vec{x}_q$ denote the corresponding sets of $x_i$ variables. Then for any fixed choice of $\vec{a}_p, \vec{x}_p$, $\vec{x}_q$, and $\vec{x}_q' \neq \vec{x}_q$, we can prove

$$\sum_{\vec{a}_q} R\left(\vec{a}_p,\vec{a}_q|\vec{x}_p,\vec{x}_q\right) = \sum_{\vec{a}_q} R\left(\vec{a}_p,\vec{a}_q|\vec{x}_p,\vec{x}_q'\right). \tag{3}$$

The condition above applies to any partition of the parties into two sets. The proof of (3), which we write out in appendix A, amounts to iterated applications of (2). Condition (3) can be equivalently re-written a little more compactly in terms of the marginal distribution of $\vec{a}_p$ as

$$R\left(\vec{a}_p|\vec{x}_p,\vec{x}_q\right) = R\left(\vec{a}_p|\vec{x}_p,\vec{x}'_q\right). \tag{4}$$

With this in mind, we can define the probability distribution $R(\vec{a}_p|\vec{x}_p)$ as

$$R\left(\vec{a}_p|\vec{x}_p\right) := R\left(\vec{a}_p|\vec{x}_p,\vec{x}_q\right) \tag{5}$$

for some fixed choice of $\vec{x}_q$—any choice of which will do, as (4) ensures there is no ambiguity in leaving this choice arbitrary. We note that if a distribution over $\vec{X}$ is presumed, so that a joint distribution of all random variables $\vec{A},\vec{X}$ exists and the conditional probabilities $R(\vec{a}_p|\vec{x}_p)$ can be calculated directly from it, satisfaction of (2) ensures that these calculations will be consistent with (5); see appendix A for a demonstration. Finally, and unsurprisingly, the reduced distribution (5) is itself nonsignaling: Subdividing $\vec{a}_p$ into two strings $\vec{a}_{p_r}$ (receiver) and $\vec{a}_{p_s}$ (signaler), we have:

$$\begin{aligned}\sum_{\vec{a}_{p_s}} R\left(\vec{a}_{p_r},\vec{a}_{p_s}|\vec{x}_{p_r},\vec{x}_{p_s}\right) &= R\left(\vec{a}_{p_r}|\vec{x}_{p_r},\vec{x}_{p_s}\right)\\ &= R\left(\vec{a}_{p_r}|\vec{x}_{p_r},\vec{x}_{p_s},\vec{x}_q\right)\\ &= R\left(\vec{a}_{p_r}|\vec{x}_{p_r},\vec{x}'_{p_s},\vec{x}'_q\right)\\ &= R\left(\vec{a}_{p_r}|\vec{x}_{p_r},\vec{x}'_{p_s}\right) \qquad = \sum_{\vec{a}_{p_s}} R\left(\vec{a}_{p_r},\vec{a}_{p_s}|\vec{x}_{p_r},\vec{x}'_{p_s}\right)\end{aligned} \tag{6}$$

where we applied (5), then (4), then (5) after converting the sums into equivalent expressions about marginal probabilities. The equality $R(\vec{a}_{p_r}|\vec{x}_{p_r},\vec{x}_{p_s}) = R(\vec{a}_{p_r}|\vec{x}_{p_r},\vec{x}'_{p_s})$ can be given an operational interpretation as the broadest notion of nonsignaling, encapsulating the idea that no subset of parties (those corresponding to $\vec{a}_{p_s}$) can signal to any disjoint other subset (those corresponding to $\vec{a}_{p_r}$) without explicit reference to the third uninvolved subset of remaining parties (those corresponding to $\vec{a}_q$).

The final property that we derive, required for some arguments in the next section, is as follows: the input-conditional distribution of a subset of parties, conditioned additionally on the inputs *and outputs* of the other parties, is nonsignaling. That is, assume a particular set of outputs $\vec{a}_q$ of the last $q$ parties occurs with nonzero probability, given the input vector $\vec{x}_q$: $R(\vec{a}_q|\vec{x}_q) > 0$. Then the natural definition of the distribution of the first $p$ parties' output $\vec{a}_p$ conditioned on their input $\vec{x}_p$, as well as the inputs $\vec{x}_q$ and outputs $\vec{a}_q$ of the $q$ group, is

$$R^{\vec{a}_q,\vec{x}_q}\left(\vec{a}_p|\vec{x}_p\right) = R\left(\vec{a}_p|\vec{x}_p,\vec{x}_q,\vec{a}_q\right) := \frac{R\left(\vec{a}_p,\vec{a}_q|\vec{x}_p,\vec{x}_q\right)}{R\left(\vec{a}_q|\vec{x}_q\right)} \tag{7}$$

where we exploit (5) to justify writing $R(\vec{a}_q|\vec{x}_q)$ instead of the more generally valid $R(\vec{a}_q|\vec{x}_p,\vec{x}_q)$ in the denominator above. It is immediate that $R^{\vec{a}_q,\vec{x}_q}(\vec{a}_p|\vec{x}_p)$ is a valid probability distribution (nonnegative and sums to one over $\vec{a}_p$). Furthermore $R^{\vec{a}_q,\vec{x}_q}(\vec{a}_p|\vec{x}_p)$ is no-signaling as follows:

$$\begin{aligned}\sum_{a_p} R\left(\vec{a}_{p-1},a_p|\vec{x}_{p-1},x_p,\vec{a}_q,\vec{x}_q\right) &= \frac{\sum_{a_p} R\left(\vec{a}_{p-1},a_p,\vec{a}_q|\vec{x}_{p-1},x_p,\vec{x}_q\right)}{R\left(\vec{a}_q|\vec{x}_q\right)}\\ &= \frac{\sum_{a_p} R\left(\vec{a}_{p-1},a_p,\vec{a}_q|\vec{x}_{p-1},x'_p,\vec{x}_q\right)}{R\left(\vec{a}_q|\vec{x}_q\right)}\\ &= \sum_{a_p} R\left(\vec{a}_{p-1},a_p|\vec{x}_{p-1},x'_p,\vec{a}_q,\vec{x}_q\right)\end{aligned} \tag{8}$$

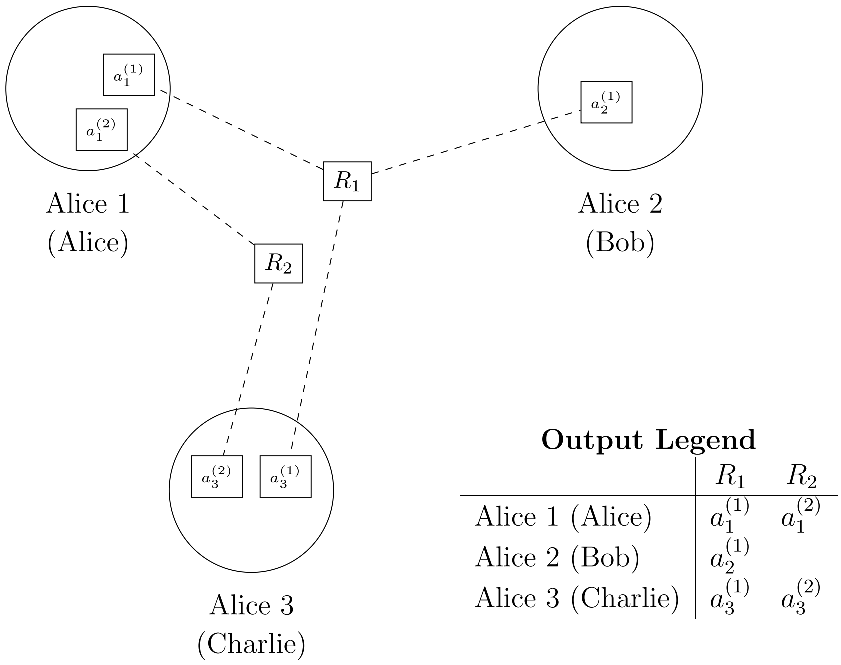


**Figure 2.** An example of three parties sharing two nonsignaling resources.

where in the middle equality we apply (2) to the numerator. As (8) is equivalent to (2) when applied to the reduced distribution $R^{\vec{a}_q,\vec{x}_q}(\vec{a}_p|\vec{x}_p)$, it naturally follows that $R^{\vec{a}_q,\vec{x}_q}(\vec{a}_p|\vec{x}_p)$ satisfies all of the properties (3)–(6) derived as a consequence of (2). Furthermore it is straightforward to confirm that conditioning as in (7) iteratively is equivalent to performing the steps all at once; i.e. if $\vec{a} = (\vec{a}_p, \vec{a}_q, \vec{a}_r)$ and we take the no-signaling distribution $R^{\vec{a}_r,\vec{x}_r}(\vec{a}_p, \vec{a}_q|\vec{x}_p, \vec{x}_q)$ and condition on the input-output combination $\vec{a}_q, \vec{x}_q$ to get $(R^{\vec{a}_r,\vec{x}_r})^{\vec{a}_q,\vec{x}_q}(\vec{a}_p|\vec{x}_p)$, the result is equivalent to $R^{\vec{a}_q\vec{a}_r,\vec{x}_q\vec{x}_r}(\vec{a}_p|\vec{x}_p)$.

### 2.2. *Networked collections of nonsignaling resources: paths and decision trees*

Having defined the nonsignaling condition (2) and derived some of its consequences, we are ready to study networked collections of nonsignaling resources. We consider a scenario of $n$ spatially separated measuring parties, which we call Alice 1 through Alice $n$ (or Alice-Bob-Charlie in scenarios of $n = 3$ parties). The $n$ parties share a set of $m$ nonsignaling resources $\mathcal{R} = \{R_1, \ldots, R_m\}$; figure 2 gives a schematic example of $n = 3$ parties sharing $m = 2$ resources. Each nonsignaling resource $R_k$ is shared by a subset of parties indexed by a set $\mathcal{M}_k = \{k_1, \ldots, k_{n_k}\} \subseteq \{1, \ldots, n\}$ whose cardinality $n_k$ can be as small as 1 and as large as $n$; in the example figure 2 these sets correspond to columns in the 'Output Legend' of figure 2, so the set $\mathcal{M}_1$ is all three parties while $\mathcal{M}_2$ is only Alice and Charlie. Each party sharing the resource $R_k$ has an *input* $X_{k_j}^{(k)}$ that can take one or more values $x_{k_j}^{(k)}$, for which there is then a corresponding *output* $A_{k_j}^{(k)}$ taking values $a_{k_j}^{(k)}$. Below, we will omit the superscripts $(k)$ from the $X$ and $A$ variables when it is clear from context to which $R_k$ they are associated. We assume that the output space for a fixed $A_p^{(k)}$ is the same for every choice of input $x_p^{(k)}$. (This is not restrictive, because we can always make it true of a resource by artificially augmenting value spaces, assigning probability zero to the added outputs; (2) will hold of the augmented distribution.)

We want to examine the joint probability distributions that arise when each party can measure the portion of the resources they share in different orders and use outputs of earlier resources to influence choices of inputs provided to later resources, as well as the order in which they access the later resources. Mathematically: letting $\vec{A}_k$ denote the vector of all outputs of resource $k$ possessed by the subset of parties $\mathcal{M}_k$, we are interested in the distribution of joint outputs of all $n$ resources $\mathbb{P}(\vec{A}_1,\ldots,\vec{A}_m|\mathbf{X}_1,\ldots,\mathbf{X}_n)$ when each party follows such a scheme. A fixed party $p$ observes one entry from each resource-output vector $\vec{A}_k$ that corresponds to a resource $R_k$ they share; $\mathbf{X}_p$ denotes an initial setting provided to the $p$th party, on which they can condition their strategy for accessing the resources. Recall that we call $\mathbf{X}_p$ the *setting* to differentiate it from the various $X_p^{(k)}$ supplied to the $R_k$ resources which we call *inputs*.

We model the scheme with a decision tree for each party $p$, which encompasses their strategy for accessing their resources: which resource $R_k$ she will access first, then how she will proceed to the next resource depending on the observed output, and so forth. After carefully stipulating the structure of these decision trees, we provide a formula for computing $\mathbb{P}(\vec{A}_1,\ldots,\vec{A}_m|\mathbf{X}_1,\ldots,\mathbf{X}_n)$, and show that the procedure is sound (i.e. leads to a well-posed probability distribution).

Given a party Alice $p$ who possesses a share of $m_p \leqslant m$ of the resources in $\mathcal{R} = \{R_1,\ldots,R_m\}$, let us denote the index set of the resources she has access to as $\mathcal{R}_p$. (Here it is visually useful to note that $\mathcal{R}_p$ corresponds to *rows* of the output legend in figure 2; in contrast the sets $\mathcal{M}_k$ defined earlier correspond to *columns*.) Then we define a decision tree as follows, with figure 3 providing an illustrative example:

**Definition.** A *decision tree* for Alice $p$ is a tree graph, consisting of nodes connected by edges, where all maximal length paths (those starting at the root node and ending at a terminal node) are of the same length, exactly $m_p + 1$ edges—the number of resources shared by party $p$, plus one. Furthermore all edges and nodes except for terminal nodes and the root node are labeled, satisfying the following conditions:

(i) There is exactly one edge leaving the root node for each choice of setting $\mathbf{x}_p$, which is labeled with this setting choice. (This tells Alice $p$ what to do for each choice of setting $\mathbf{X}_p$).

(ii) Every non-root and non-terminal node is labeled with two entries instructing Alice $p$ what to do. If $i$ is the number of edges downstream from the root node—this is known as the depth, or **level**, of the node—we notate these two values $(c_i, inp_i)$, where $c_i$ represents the choice of resource to use, and $inp_i$ represents the choice of input to provide to the corresponding resource $R_{c_i}$. The label $c_i$ is never equal to the earlier $c_j$ label appearing in one of its ancestor nodes—a resource is only used once.

(iii) For every node carrying a $(c_i, inp_i)$ label (i.e. non-root, non-terminal nodes), the number of edges descending from it is equal to the number of valid outputs from the resource $R_{c_i}$ for party $p$'s output $A_p^{(k)}$. Each emerging edge is labeled with a unique one of these valid outputs, which we notate $out_i$; the subtree descending from this edge represents what the party proceeds to do conditioned on observing this particular output.

Figure 3 is an example of a decision tree for Alice 1 in the three-party, two-shared-resources experiment of figure 2. The decision tree framework can be slightly augmented if we want to have each party $p$ report an overall outcome upon reaching a terminal node, depending on

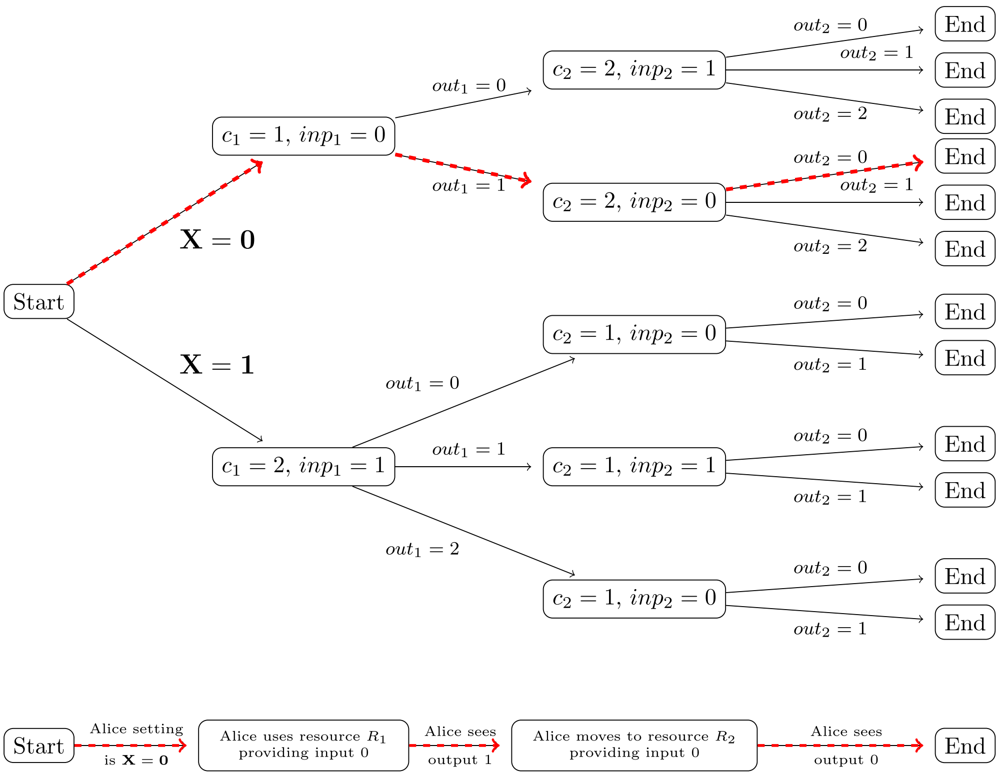


**Figure 3.** A decision tree. The following example is for Alice 1 in the 3-party, 2-resource scenario of figure 2 where Alice 1 shares resource $R_1$ with Alice 2 (Bob) and Alice 3 (Charlie), and shares resource $R_2$ with Charlie only. Her setting $\mathbf{X}$ can take one of two values $\{\mathbf{0},\mathbf{1}\}$; note that which resource she consults first depends on this setting. Alice can observe one of two possible outputs $\{0,1\}$ from resource $R_1$ and three possible outputs $\{0,1,2\}$ from resource $A_2$. A sample path is highlighted with dashes; Alice's actions and observations for this sample path are detailed at the bottom of the figure.

which terminal node is reached: we model this as the 'final outcome' $\mathbf{A}_p$ where disjoint subsets of terminal nodes are labeled with different values $\mathbf{a}_p$.

Note that in condition (iii), it is possible for a valid output to occur with probability zero; these are still included on the tree just to avoid some cluttering caveats in the formal arguments below. In a similar vein, conditions (i)–(iii) ensure that in every maximal-length path connecting the root node to a terminal node, the sequence $c_1,\ldots,c_{m_p}$ contained in the traversed nodes maps bijectively to $\mathcal{R}_p$ (every resource is consulted exactly once). The framework above thus assumes that all parties always use every resource they have access to. This assumption makes the construction of the joint distribution in the next section a little more natural, and is not actually restrictive—we discuss how to account for the possibility of 'unused' resources further below.

The key observation about the structure of a party $p$'s decision tree, which enables the sound construction of the joint probability distribution, is as follows: given a setting $\mathbf{x}_p$, a fixed choice of outputs $a_p^{(k)}$ for each resource $R_k$ shared by party $p$ uniquely determines a max-length path through the decision tree. This is can be confirmed visually by following through figure 3: If we are told the setting $\mathbf{X}_1$ is $\mathbf{0}$, then the assignment $A_1^{(1)} = 1$ and $A_1^{(2)} = 0$ uniquely corresponds to the highlighted path in the tree. Any other assignment $A_1^{(1)} = a_1^{(1)}$ and $A_1^{(2)} = a_1^{(2)}$ corresponds to a different unique branch. Once the branch of the tree is determined, the $inp_i$ labels contained in the nodes along this path specify the inputs that the party must have provided to each resource. Thus each resource input $x_p^{(k)}$ is a function of the string of $a_p^{(k)}$ (as well as the initial setting $\mathbf{x}_p$), and so could be written $x_p^{(k)}\left(\mathbf{x}_p, a_{k_1}^{(k)}, \ldots, a_{k_{n_k}}^{(k)}\right)$, though we do not use this explicit functional notation below. Indeed, the soundness of the joint probability distribution construction will depend crucially on a further observation that each $x_p^{(k)}$ is determined by only a *proper subset* of the $a_p^{(k)}$: if we work down the initial segment of a path descending to level $i$, this initial segment is determined by $\mathbf{X}_p$ and $i-1$ choices of $A_p^k$, and fixes $i$ choices of $x_p^{(k)}$ that will then be constant independent of the other $A_p^{(k)}$.

For example: consider a party who possesses a share of five resources $R_1, R_2, R_3, R_4, R_5$, and assume a setting choice $\mathbf{X}_p = \mathbf{x}_p$. Consider a fixed string of outputs $a_p^{(1)}, a_p^{(2)}, a_p^{(3)}, a_p^{(4)}, a_p^{(5)}$. Then it is true that the only way that Alice $p$ ends up having observed this specific choice of $a_p^{(1)}, a_p^{(2)}, a_p^{(3)}, a_p^{(4)}, a_p^{(5)}$ is to have traversed a specific unique path through the decision tree, in which she observed these outputs in a certain order and provided specific resource inputs along the way. Suppose that on this path, we have $c_1 = 3, c_2 = 2, c_3 = 5$, so that Alice $p$ must have initially consulted resource $R_3$, then $R_2$, then $R_5$ to be consistent with observing the given output string. Then for other strings of outputs $A_p^{(1)}, a_p^{(2)}, a_p^{(3)}, A_p^{(4)}, a_p^{(5)}$ with any different values of $A_p^{(1)}$ and $A_p^{(4)}$, the corresponding path on the decision tree will have a same initial segment, and map to the same resource input choices of $x_p^{(2)}$, $x_p^{(3)}$, and $x_p^{(5)}$ (and indeed one additional $x_p^{(k)}$ will be determined by $inp_4$, where $k = c_4$). Notice that alternate choices of $A_p^{(2)}, A_p^{(3)}, A_p^{(5)}$ do not necessarily determine $X_p^{(2)}$, $X_p^{(3)}$, and $X_p^{(5)}$: it could be on the decision tree that if $A_p^{(2)}$ is equal to a different $a_p'^{(2)}$, then $c_3$ equals (say) 4 instead of 5 so that $R_4$ is the third resource used; then, we would have instead all strings of the form $A_p^{(1)}, a_p'^{(2)}, a_p^{(3)}, a_p^{(4)}, A_p^{(5)}$ determine the same $x_p^{(2)}, x_p^{(3)}, x_p^{(4)}$ (and one additional $x_p^{(k)}$ determined by $inp_4$ and $c_4$) for all choices of $A_p^{(1)}$ and $A_p^{(5)}$.

Before moving to the construction of the joint probability distribution, let us briefly return to the question of modeling situations where a party might not use a resource, or may decide to use it only conditionally on seeing certain outputs from other resources. Within the above framework, we can model this a couple of different ways. One option is to introduce an input choice $\perp$ intended to mean 'unused:' if party $q$ supplies the input $\perp$ for $X_q$, the resource then with probability 1 returns for $A_q$ a 'no output recorded' result which we also denote as $\perp$; the distribution for the non-$q$ parties $R^{\perp_q \perp_q}(\cdots|\cdots)$ is then just their marginal (5). Another option that avoids the introduction of an extra input choice is to collect all unused resources and put them at the end of the decision tree with an arbitrary dummy choices of input provided, where all outputs lead to the same subtree—the output is essentially ignored, as the party behaves the same way no matter the output value.

We also note that local probabilistic choices can be encompassed by our framework: if, for instance, Alice $p$ at some point decides to flip a fair coin and condition her input to a later resource based on the coin result, we can model this coin flip as a one-party, single-choice-of-input resource $R_k(A_p|X_p)$ satisfying $R_k(A_p|X_p = \text{the one input}) = 1/2$. As the degenerate input plays no role, we omit it and represent such resources as $R_k(A_p)$. Multiple parties can also share such input-free resources, which will look like (for example) $R_k(A_2, A_3)$. Such resources correspond to *shared local randomness* and have an operational interpretation as a random process whose output is distributed to the parties prior to the beginning of the experiment. As is well known in the literature, it is convenient to model such choices by combining all of them into a single classical random resource that is shared by all parties; this can be encompassed in our framework and we will return to it more formally later.

We now move on to building the joint distribution from the decision trees of each party, and showing that it is consistent/well-defined (i.e. resulting in a normalized probability distribution).

### *2.3. Determining the joint distribution $\mathbb{P}(\vec{A}_1,\ldots,\vec{A}_m|\mathbf{X}_1,\ldots,\mathbf{X}_n)$. Soundness of the method*

Given a candidate probability $\mathbb{P}(\vec{a}_1,\ldots,\vec{a}_m|\mathbf{x}_1,\ldots,\mathbf{x}_n)$ for a fixed choice of $\vec{a}_1,\ldots,\vec{a}_m$ and $\mathbf{x}_1,\ldots,\mathbf{x}_n$, we assign a value between 0 and 1 according to the following method: For each party $p\in\{1,\ldots,n\}$, locate the unique branch (maximal-length path) from their decision tree determined by $\mathbf{x}_p$ and that party's $a_p^{(k)}$ outputs extracted from among the $\vec{a}_1,\ldots,\vec{a}_m$. Then note the resource inputs $x_j^{(k)}$ that are determined by the $inp_i$ along these paths, and set

$$\mathbb{P}(\vec{a}_1,\ldots,\vec{a}_m|\mathbf{x}_1,\ldots,\mathbf{x}_n)=\prod_{k=1}^{m}R_k(\vec{a}_k|\vec{x}_k)=\prod_{k=1}^{m}R_k\left(a_{k_1}^{(k)},\ldots,a_{k_{n_k}}^{(k)}|x_{k_1}^{(k)},\ldots,x_{k_{n_k}}^{(k)}\right)\tag{9}$$

where we have written out $\vec{a}_k$ as $a_{k_1}^{(k)},\ldots,a_{k_{n_k}}^{(k)}$ on the right. The product form of (9) reflects the intuitive notion that the different resources are indeed different and so operate independently of each other; this eventually underpins the derivation of nontrivial constraints in paradigms such as LOSR-GMNL [12]. The sense in which the resources 'operate independently' is not quite the same as independence of random variables/events with the attendant standard factorization rule $\mathbb{P}(S\cap T)=\mathbb{P}(S)(T)$: a more relevant (though signaling) analogy would be a scenario of two telephones whose inner workings are completely separate (so 'independent') but one can take what one hears from one telephone (output) and repeat it into the other (as input). Thus in (9), while the conditional distributions factor, for a party $p$ an input $x_p^{(k)}$ to one resource $R_k$ can depend on (be a function of) an output $a_p^{(k')}$ from another resource $R_{k'}$, with the form of the dependence dictated by party $p$'s decision tree. We remark that this notion of independence of resources is important in related but different approaches such as the study of network nonlocality [29].

As an example to illustrate how (9) is computed, consider the three-party scenario of figure 2. Here we have

$$\begin{aligned}\mathbb{P}(\vec{a}_1,\vec{a}_2|\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)&=\mathbb{P}\left(\overbrace{a_1^{(1)},a_2^{(1)},a_3^{(1)}}^{\vec{a}_1},\overbrace{a_1^{(2)},a_3^{(2)}}^{\vec{a}_2}|\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3\right)\\&=R_1\left(a_1^{(1)},a_2^{(1)},a_3^{(1)}|x_1^{(1)},x_2^{(1)},x_3^{(1)}\right)R_2\left(a_1^{(2)},a_3^{(2)}|x_1^{(2)},x_3^{(2)}\right)\\&=R_1(a_1,a_2,a_3|x_1,x_2,x_3)R_2(a_1,a_3|x_1,x_3),\end{aligned}\tag{10}$$

where we remove the $(k)$ superscripts in the last line as they are redundant within an $R_k(\cdots|\cdots)$ expression. If party 1's decision tree is as in figure 3, and if on the left side of (10) we have $\mathbf{x}_1$, $a_1^{(1)}$, and $a_1^{(2)}$ as $\mathbf{0}$, 1, and 0 respectively, then this corresponds to the highlighted path in the figure. The $inp_i$ along this path then determine the corresponding $x_1^{(1)}$ and $x_1^{(2)}$ values that will appear on the right side of (10); specifically, we then obtain $R_1(1,a_2,a_3|0,x_2,x_3)R_2(0,a_3|0,x_3)$.

To complete the computation of the probability, one would then consult decision trees for the second and third parties to fill in the remaining values.

We claim (9) yields a valid joint probability distribution for each choice of $\mathbf{x}_1,\ldots,\mathbf{x}_n$. Nonnegativity is immediate: the right side of (9) is a product of nonnegative terms. Showing that the sum over all values of $\vec{a}_1,\ldots,\vec{a}_m$ is equal to 1 is more involved, and depends critically on the fact that the $R_k$ resources are nonsignaling.

To further motivate our derivations below, let us discuss the potential *failure* of normalization if the resources are signaling. Figure 1 in [15] leads to such a failure: here, there are two parties sharing two signaling resources that they access in opposite order resulting in a sort of 'grandfather paradox' inconsistency as described in further detail in that paper. Mathematically, if we try to apply our formula (9) to this example, the right hand side will always take the form $R_1(a_1,\beta|\delta,b_2)R_2(\alpha,b_2|a_1,\gamma)$ with the multiple appearances of $a_1$ and $b_2$ resulting from Alice consulting $R_1$ first and using her output as input to $R_2$, while Bob consults $R_2$ first and uses his output as input to $R_1$. Then with the signaling properties of the $R_1$ and $R_2$ distributions as described in [15], at least one of the resources $R_1$ and $R_2$ assigns zero probability for all choices of $a_1$ and $b_2$ independently of the other entries $\alpha,\beta,\gamma,\delta$, so all probabilities assigned by (9) are zero and thus cannot sum to one. We do not encounter such problems when the nonsignaling condition (2) is satisfied by the resources.

The proof of normalization, while not immediate, is also not exceedingly involved. However, applying it directly to the general equation (9) requires some unwieldy notation that can obfuscate what is going on. Hence we first illustrate the key idea with the three-party, two-resource example of figure 2. Summing (9) over all outputs will yield

$$\sum_{\vec{a}_1,\vec{a}_2}\mathbb{P}\left(\vec{a}_1,\vec{a}_2|\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3\right)=\sum_{a_1^{(1)},a_2^{(1)},a_3^{(1)},a_1^{(2)},a_3^{(2)}}R_1\left(a_1,a_2,a_3|x_1,x_2,x_3\right)R_2\left(a_1,a_3|x_1,x_3\right),\tag{11}$$

where we can re-write the right side above in a little more readable fashion, using a standard Alice-Bob-Charlie renaming, as

$$\sum_{a^{(1)},b^{(1)},c^{(1)},a^{(2)},c^{(2)}}R_1\left(abc|xyz\right)R_2\left(ac|xz\right).\tag{12}$$

Now, certain $x^{(k)}$, $y^{(k)}$ and $z^{(k)}$ values can depend on $\vec{a}_1$ and $\vec{a}_2$ and thus can vary as the sum is performed. However, the input to a party's *first* used resource depends only on their setting $\mathbf{x}_p$. Let us suppose that for the given choices of $\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3$, Alice's and Bob's first steps in their decision trees are to consult resource $R_1$, whereas Charlie consults the other resource $R_2$. Then it is only $z^{(1)}$ and $x^{(2)}$ that can vary in (12). In particular, $z^{(2)}$ is constant in the sum which will allow us to pull a term out as follows. First, re-write (12) as

$$\begin{aligned}&\sum_{c^{(2)}}\sum_{a^{(1)},b^{(1)},c^{(1)},a^{(2)}}R_1\left(abc|xyz\right)R_2\left(ac|xz\right)\\&\quad=\sum_{c^{(2)}:R_2(c|z)>0}\sum_{a^{(1)},b^{(1)},c^{(1)},a^{(2)}}R_1\left(abc|xyz\right)R_2\left(ac|xz\right),\end{aligned}$$

where the restriction of the outer sum is valid because given any choice of $c^{(2)}$ for which $R_2(c|z) = 0$ holds, $R_2(ac|xz)$ will equal zero as well, and hence the corresponding term in the summand is zero. Then, by (7) we can make the substitution

$$R_2\left(ac|xz\right) = R_2\left(c|z\right) R_2\left(a|xz,c\right) \tag{13}$$

and pull out $R_2(c|z)$ to write

$$= \sum_{c^{(2)}:R_2(c|z)>0} R_2\left(c|z\right) \sum_{a^{(1)},b^{(1)},c^{(1)},a^{(2)}} R_1\left(abc|xyz\right) R_2\left(a|xz,c\right). \tag{14}$$

We remark that this step may fail in the absence of the no-signaling assumption; one would be attempting to pull out $R_2(c|xz)$ instead of just $R_2(c|z)$, and Alice's input $x^{(2)}$ to $R_2$ might not be independent of her output $a^{(1)}$ from $R_1$ (which she consulted first). Continuing on, we can enlist the fact that $x^{(1)}$ and $y^{(1)}$ similarly do not vary in the sum, so that the process can be repeated on the inner sum to rewrite (14) as

$$\sum_{c^{(2)}:R_2(c|z)>0} R_2\left(c|z\right) \sum_{a^{(1)},b^{(1)}:R_1(ab|xy)>0} R_1\left(ab|xy\right) \sum_{c^{(1)},a^{(2)}} R_1\left(c|xyz,ab\right) R_2\left(a|xz,c\right). \tag{15}$$

We have pulled out the probabilities corresponding to the first resource each party consults.

Now looking at $R_1(c|xyz,ab)$ in the innermost sum, the inputs $x^{(1)}$ and $y^{(1)}$ are fixed, but Charlie's input $z^{(1)}$ can depend on his output $c^{(2)}$ from $R_2$ which he consulted earlier on his decision tree. However, for each choice of $c^{(2)}$ in the outermost sum, $z^{(1)}$ *is* fixed; and for each fixed choice of $a^{(1)}$ and $b^{(1)}$ in the middle sums, $R_1(c|xyz,ab)$ will be a single probability distribution for which we are summing over all outputs $c^{(1)}$ in the innermost sum. We can make a parallel argument for $R_2(a|xz,c)$. So for each fixed choice of $a^{(1)}, b^{(1)}, c^{(2)}$ the inner sum is

$$\sum_{c^{(1)},a^{(2)}} R_1\left(c|xyz,ab\right) R_2\left(a|xz,c\right) = \sum_{c^{(1)}} R_1\left(c|xyz,ab\right) \sum_{a^{(2)}} R_2\left(a|xz,c\right) = 1.$$

Then (15) reduces to just the outer sums, for which we have

$$\sum_{c^{(2)}:R_2(c|z)>0} R_2\left(c|z\right) \sum_{a^{(1)}b^{(1)}:R_1(ab|xy)>0} R_1\left(ab|xy\right) = 1.$$

The above example contains the essence of the proof of normalization. For scenarios involving decision trees of depth 3 or more the process of replacing (12) with (15) must be applied iteratively to the inner sum in (15): pulling out probabilities corresponding to the second consulted resource on a decision tree, then the third, etc. There is a slight notational complication because the choice of *which* resource is consulted next may change depending on the values fixed by outer sums (i.e. the outputs of the previously consulted resource). For completeness we present the proof of the general case in appendix B.

## 3. Properties of the induced distribution

The distribution defined by (9) satisfies a number of important properties that we describe here.

### 3.1. Nonsignaling of the induced distribution

We can demonstrate that the distribution $\mathbb{P}$ defined in (9) is nonsignaling, in the following sense illustrated for our 3-party example. Recall that for this example $\mathbb{P}$ is given by

$$\mathbb{P}\left(\vec{a}_1, \vec{a}_2 | \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\right) = R_1\left(a_1, a_2, a_3 | x_1, x_2, x_3\right) R_2\left(a_1, a_3 | x_1, x_3\right).$$

Now re-ordering the outputs in the front of $\mathbb{P}(\cdots|\cdots)$ so that they are grouped by party instead of by resource, we can re-write the probability as

$$\mathbb{P}\left(\underbrace{a_1^{(1)}, a_1^{(2)}}_{\mathbf{a}_1}, \underbrace{a_2^{(1)}}_{\mathbf{a}_2}, \underbrace{a_3^{(1)}, a_3^{(2)}}_{\mathbf{a}_3} | \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\right).$$

Then, the assertion is that $\mathbb{P}(\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3 | \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ satisfies the no-signaling condition (2): $\sum_{\mathbf{a}_1} \mathbb{P}(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3 | \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = \sum_{\mathbf{a}_1} \mathbb{P}(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3 | \mathbf{x'}_1, \mathbf{x}_2, \mathbf{x}_3)$ for all fixed choices of the non-summed over variables, and the corresponding equalities hold for the parallel expressions with $\sum_{\mathbf{a}_2}$ and $\sum_{\mathbf{a}_3}$. Here, we are taking the 'final outcome' of party $p$—discussed after the definition of decision trees in section 2.2—to be the complete transcript of all resource outputs recorded by that party. If party $p$ instead bins together some of these transcripts to report a final outcome as some non-injective function of the complete transcript of resource outputs, the nonsignaling property of the distribution of these binned final outcomes will follow from the nonsignaling property of the distribution of complete transcripts.

To prove the nonsignaling property in the general $n$-party, $m$-resource setting, recall $\mathcal{R}_p \subseteq \{1, \ldots, m\}$ denotes the subset of $k$ indices that correspond to resources $R_k$ that are shared by party $p$; these correspond to rows in the output legend of figure 2. Then we want to show that for each fixed choice of $p$,

$$\sum_{a_p^{(k)} : k \in \mathcal{R}_p} \mathbb{P}\left(\vec{a}_1, \ldots., \vec{a}_m | \mathbf{x}_1, \ldots, \mathbf{x}_p, \ldots, \mathbf{x}_n\right) = \sum_{a_p^{(k)} : k \in \mathcal{R}_p} \mathbb{P}\left(\vec{a}_1, \ldots., \vec{a}_m | \mathbf{x}_1, \ldots, \mathbf{x'}_p, \ldots, \mathbf{x}_n\right) \tag{16}$$

for all fixed choices of the non-summed-over variables and settings. We prove this as follows for $p = 1$ (the proof applies without loss of generality to the other parties). First, let us deal with a trivial case: suppose that for some resource $k$ shared by party 1, we have $R_k(a_{k_2}, \ldots, a_{k_{n_k}} | x_{k_2}, \ldots, x_{k_{n_k}}) = 0$; that is, the marginal probability of the other parties' outputs is zero. Enlisting (5), this implies that for any choice of $x_1$,

$$0 = R_k\left(a_{k_2}, \ldots, a_{k_{n_k}} | x_1, x_{k_2}, \ldots, x_{k_{n_k}}\right) = \sum_{a_1} R_k\left(a_1, a_{k_2}, \ldots, a_{k_{n_k}} | x_1, x_{k_2}, \ldots, x_{k_{n_k}}\right)$$

and so $R_k(a_1, a_{k_2}, \ldots, a_{k_{n_k}} | x_1, x_{k_2}, \ldots, x_{k_{n_k}})$ must equal zero for all choices of $a_1$. This implies both sides of the equality in (16) are zero, after re-expressing $\mathbb{P}(\cdots|\cdots)$ terms as products of $R(\cdots|\cdots)$ terms according to (9). So let us now assume that all $R_k(a_{k_2}, \ldots, a_{k_{n_k}} | x_{k_2}, \ldots, x_{k_{n_k}})$ are positive. Now using (9) to write the left side of (16) as a product of $R(\cdots|\cdots)$ terms, factoring out those not shared by party 1, and applying (7), we can write

$$\begin{aligned}
&\sum_{a_1^{(k)}:k\in\mathcal{R}_1} \mathbb{P}(\vec{a}_1,\ldots,\vec{a}_m|\mathbf{x}_1,\ldots,\mathbf{x}_n) \\
&= \sum_{a_1^{(k)}:k\in\mathcal{R}_1} \prod_{k=1}^{m} R_k\left(a_{k_1},\ldots,a_{k_{n_k}}|x_{k_1},\ldots,x_{k_{n_k}}\right) \\
&= \prod_{k\notin\mathcal{R}_1} R_k(\cdots|\cdots) \sum_{a_1^{(k)}:k\in\mathcal{R}_1} \prod_{k\in\mathcal{R}_1} R_k\left(\underbrace{a_{k_1}}_{a_1^{(k)}},a_{k_2},\ldots,a_{k_{n_k}}|\underbrace{x_{k_1}}_{x_1^{(k)}},x_{k_2},\ldots,x_{k_{n_k}}\right) \\
&= \prod_{k\notin\mathcal{R}_1} R_k(\cdots|\cdots) \sum_{a_1^{(k)}:k\in\mathcal{R}_1} \prod_{k\in\mathcal{R}_1} R_k\left(a_{k_2},\ldots,a_{k_{n_k}}|x_{k_2},\ldots,x_{k_{n_k}}\right) R_k\left(a_1|x_1,x_{k_2},\ldots,x_{k_{n_k}},a_{k_2},\ldots,a_{k_{n_k}}\right) \\
&= \prod_{k\notin\mathcal{R}_1} R_k(\cdots|\cdots) \prod_{k\in\mathcal{R}_1} R_k\left(a_{k_2},\ldots,a_{k_{n_k}}|x_{k_2},\ldots,x_{k_{n_k}}\right) \sum_{a_1^{(k)}:k\in\mathcal{R}_1} \prod_{k\in\mathcal{R}_1} R_k\left(a_1|x_1,x_{k_2},\ldots,x_{k_{n_k}},a_{k_2},\ldots,a_{k_{n_k}}\right) \\
&= \prod_{k\notin\mathcal{R}_1} R_k(\cdots|\cdots) \prod_{k\in\mathcal{R}_1} R_k\left(a_{k_2},\ldots,a_{k_{n_k}}|x_{k_2},\ldots,x_{k_{n_k}}\right) \qquad (17)
\end{aligned}$$

where the last equality holds because the sum in the penultimate line evaluates to one—a result that can be obtained by noting that this sum is a quantity of the form (B.4) in appendix B, and thus equals 1 by the arguments presented there[1]. Now the final expression in (17) has no variables belonging to party 1; all $R_k(a_{k_2},\ldots,a_{k_{n_k}}|x_{k_2},\ldots,x_{k_{n_k}})$ terms depend solely on the settings $\mathbf{x}_i$ and decision trees of the other parties. Since the same expression can be reached if we apply these manipulations to the right side of (16), the equality holds.

### 3.2. *Shared local randomness and local deterministic distributions*

Consider an $n$-party paradigm in which arbitrary shared local randomness is allowed: the parties are allowed to consult $n$-party no-input resources $R(a_1,\ldots,a_n)$ of arbitrary distribution. However, restrictions are imposed on the type of nonlocal resources with inputs that can be consulted. An important motivator for this paradigm is the LOSR-GMNL definition of [12], where global local shared randomness is considered a free resource always available to all $n$ parties, and it is the networks where nonlocal nonsignaling resources (with inputs) are restricted to subsets of two parties that are considered (only) bipartite nonlocal—or for a more generalized hierarchical notion of LOSR-GMNL [16], networks allowing nonlocal nonsignaling resources shared among subsets of at most $n-1$ parties are classified as *not* genuinely $n$-partite nonlocal.

If we study the class of behaviors satisfying such a paradigm, this class is equivalent to the following: convex mixtures of behaviors induced by networks comprising only extremal nonlocal nonsignaling resources satisfying whatever restrictions were previously imposed on

[1] We can alternatively obtain equality to 1 as a consequence of the normalization of joint probability distributions defined by (9): for a fixed choice of $(x_{k_2},\ldots,x_{k_{n_k}})$ and $(a_{k_2},\ldots,a_{k_{n_k}})$, $R_k(a_1|x_1,x_{k_2},\ldots,x_{k_{n_k}},a_{k_2},\ldots,a_{k_{n_k}}) = R_k^{a_{k_2},\ldots,a_{k_{n_k}},x_{k_2},\ldots,x_{k_{n_k}}}(a_1|x_1)$ can be viewed as a one-party nonsignaling resource; then, the product of these reduced $R_k^{a_{k_2},\ldots,a_{k_{n_k}},x_{k_2},\ldots,x_{k_{n_k}}}$ resources appearing in the penultimate line of (17) is the expression for the probability distribution of a one-party network of nonsignaling resources accessed according to party 1's decision tree; summing over all outputs then yields one as a consequence of normalization as proved in section 2.3.

the nonlocal resources. Here, *extremal* resources refers to resources that are extremal in the polytope of nonsignaling resources as defined and discussed in for example [10, 30]; by saying *nonlocal* we specify that these extremal behaviors are not the local deterministic (classical) ones. The equivalence is useful because it means a linear constraint on behaviors induced by networks of extremal nonlocal nonsignaling resources will automatically translate to a linear constraint on the whole class of behaviors by convexity; this technique was used in for example [11].

Let us prove the equivalence. The first key property is that shared local randomness resources like $R(a_1,a_3)$ or $R(a_1,\ldots,a_n)$ can always be factored out of the distribution entirely, in the following sense: if $R_1$ is such a resource, we can write

$$\begin{aligned}\mathbb{P}\left(\vec{a}_1,\vec{a}_2,\ldots,\vec{a}_m|\mathbf{x}_1,\ldots,\mathbf{x}_n\right) &= R_1\left(\vec{a}_1\right)\prod_{k=2}^{m} R_k\left(\vec{a}_k|,\vec{x}_k\right)\\ &= R\left(\vec{a}_1\right)\mathbb{P}_{\vec{a}_1}\left(\vec{a}_2,\ldots,\vec{a}_m|\mathbf{x}_1,\ldots,\mathbf{x}_n\right)\end{aligned} \tag{18}$$

where $\mathbb{P}_{\vec{a}_1}$ is the distribution that obtains if all parties modify their decision trees as follows: remove all consultations of the resource $R_1$; where such consultations previously occurred, instead proceed directly to the subsequent subtree that followed in the original tree when the output corresponding to $\vec{a}_1$ was observed. Figure 4 provides an illustration of this excision/bypassing procedure. Observe that the inputs $\vec{x}_k$ of the other resources will be the same whether $\mathbb{P}$ and $\mathbb{P}_{\vec{a}_1}$ are expanded according to (9), which is why (18) holds.

A key aspect of (18) is the following operational interpretation: that of a scenario where the random process $R_1$ is sampled prior to the experiment and the output $\vec{a}_1$ is distributed to all parties; then when the experiment is run, they proceed with the $\vec{a}_1$-indexed decision tree. Indeed, if there are multiple local random resources $R_1$ through $R_t$, they can all be pulled out front as

$$\begin{aligned}\mathbb{P}\left(\vec{\mathbf{A}}|\vec{\mathbf{X}}\right) &= \prod_{k=1}^{t} R_k\left(a_1,\ldots,a_n\right)\prod_{k=t+1}^{m} R_k\left(\vec{a}_k|,\vec{x}_k\right)\\ &= \prod_{k=1}^{t} R_k\left(\vec{a}_k\right)\mathbb{P}_{\vec{a}_1,\ldots,\vec{a}_t}\left(\vec{a}_{t+1},\ldots,\vec{a}_m\right),\end{aligned}$$

and $\prod_{k=1}^{t} R_k(\vec{a}_k)$ can be interpreted as a single combined shared resource $R_{\vec{k}}$ with the distribution $R_{\vec{t}}(\vec{a}_1,\ldots,\vec{a}_t) = \prod_{k=1}^{t} R_k(\vec{a}_k)$. Hence any behavior in the class $\mathbb{P}(\vec{\mathbf{A}}|\vec{\mathbf{X}})$ is equivalent to a convex mixture of behaviors induced by networks of only resources $R$ with inputs.

A further reduction can be performed. As mentioned earlier, the set of all nonsignaling resource $R(a_1,\ldots,a_n|x_1,\ldots,x_n)$ for a fixed number of parties, inputs, and outputs, comprises a *polytope*, as it is the set of behaviors satisfying linear equalities (2) along with the linear equalities and inequalities that define valid probability distributions. As such, this polytope will have a certain number $N$ of extreme points $R_i^{\text{ext}}(a_1,\ldots,a_n|x_1,\ldots,x_n)$, $i\in\{1,\ldots,N\}$, for which a general $R(a_1,\ldots,a_n|x_1,\ldots,x_n)$ can be written as a convex combination:

$$R\left(a_1,\ldots,a_n|x_1,\ldots,x_n\right) = \sum_i p\left(i\right) R_i^{\text{ext}}\left(a_1,\ldots,a_n|x_1,\ldots,x_n\right), \tag{19}$$

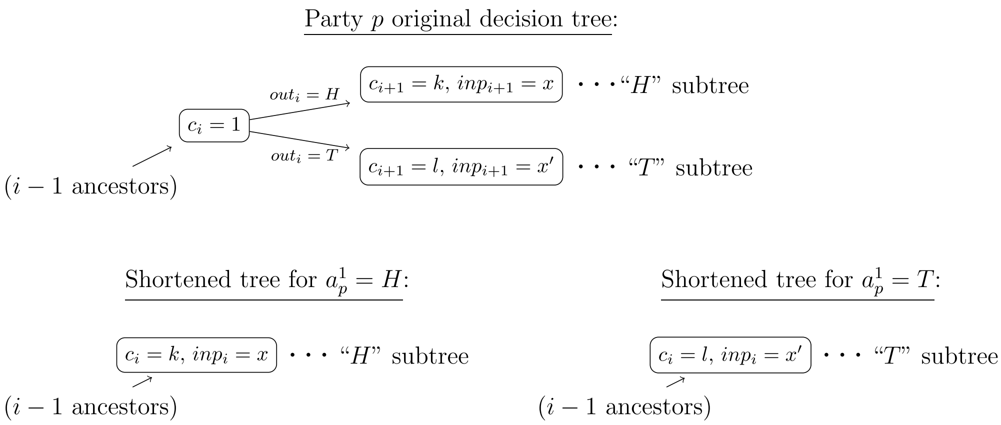


**Figure 4.** Removing local randomness from decision trees. If $R_1$ is a no-input resource, it can be thought of as shared local randomness. Such resources can be removed from decision trees as follows: for party $p$'s original decision tree (top), locate every appearance of the resource (nodes with $c_i = 1$), then create a new decision tree for each possible choice of output of the resource (here, there are two outputs) by excising the step where $R_1$ is consulted, bypassing it as though the given choice of output had been observed. Above, we create a new '$H$' decision tree by replacing every instance of the top subtree with the shortened below-left subtree, or a new '$T$' decision tree by replacing with the shortened below-right subtree.

where $p(i)$ is a probability distribution over the values of $i$. Employing such an expression for $R_1(\vec{a}_1|\vec{x}_1)$, we can write

$$
\begin{aligned}
\mathbb{P}(\vec{a}_1,\vec{a}_2,\ldots,\vec{a}_m|\mathbf{x}_1,\ldots,\mathbf{x}_n) &= R_1(\vec{a}_1|\vec{x}_1)\prod_{k=2}^{m} R_k(\vec{a}_k|,\vec{x}_k) \\
&= \left[\sum_i p(i) R_i^{\text{ext}}(\vec{a}_1|\vec{x}_1)\right]\prod_{k=2}^{m} R_k(\vec{a}_k|,\vec{x}_k) \\
&= \sum_i p(i)\left[R_i^{\text{ext}}(\vec{a}_1|\vec{x}_1)\prod_{k=2}^{m} R_k(\vec{a}_k|,\vec{x}_k\right]
\end{aligned}
\tag{20}
$$

and the term in brackets in (20) is equal to $\mathbb{P}_i(\vec{a}_1,\vec{a}_2,\ldots,\vec{a}_m|\mathbf{x}_1,\ldots,\mathbf{x}_n)$, which we define to be the distribution induced when each party uses their original decision with the single change of replacing consultations of $R_1$ with consultations of $R_i^{\text{ext}}$. Hence $\mathbb{P}$ is equal to the convex mixture $\sum_i p(i)\mathbb{P}_i$. This process can be repeated for all $R_k$ so that $\mathbb{P}$ is a convex mixture of distributions each induced by extremal-only resources.

As a final simplification, consider local deterministic distributions. These distributions are extremal in the nonsignaling polytope, but they do not exhibit nonlocal behavior: they are classical where each party's output is a deterministic function of their local input. That is, there is a function $f$ mapping each input $x_i$ to a fixed value in the range of $A_i$ for which

$$
R(a_1,\ldots,a_n|x_1,\ldots,x_n) = \prod_{i=1}^{n}\delta_{a_i,f(x_i)}
$$

where the Kronecker delta function $\delta_{a_i,f(x_i)}$ maps to one if $a_i = f(x_i)$ and zero otherwise. Such a resource can be removed from decision trees with no meaningful consequences for the induced distribution $\mathbb{P}$ by merely bypassing all steps where it is consulted: there is only one relevant edge descending from a consultation of $R$, the one labeled with the probability-one output $f(x_i)$. Specifically, for a local deterministic $R_1$ we can write $\mathbb{P}(\vec{a}_1,\vec{a}_2,\ldots,\vec{a}_m|\mathbf{x}_1,\ldots,\mathbf{x}_n) = \mathbb{P}'(\vec{a}_2,\ldots,\vec{a}_m|\mathbf{x}_1,\ldots,\mathbf{x}_n)\prod_{i=1}^{n}\delta_{a_i,f(x_i)}$ where $\mathbb{P}'$ is the distribution resulting from the shortened decision trees; $\mathbb{P}$ has no meaningful characteristics not already possessed by $\mathbb{P}'$.

### 3.3. A discussion of causality

The inflation technique [25] is an important tool for deriving constraints on behaviors that can arise in multiparty networks. It applies to all theories that are *causal*, which includes quantum mechanics, the scenario of wired nonsignaling boxes studied in this paper, and even more general probabilistic theories that would allow generalized analogs of entangled measurements on the nonsignaling resources. It is accepted in [12, 16] that the theory of wired nonsignaling boxes is causal. However, as discussed in the introduction, some references cited regarding this question [22–24] are somewhat abstract/general and focused on other questions, not explicitly addressing the matter in regards to wired nonsignaling boxes. It is thus useful to spend a few paragraphs discussing how the results derived in earlier sections imply the causality of wired nonsignaling boxes.

A definition of a causal theory amenable to the inflation technique can be found in section IIB of [16]: a theory is causal if it satisfies the conditions of Definition 1 therein along with 'device replicability.' We paraphrase the proffered definition of causality loosely as follows: consider a theory with multipartite resources (for us, the $R_k$ resources), parties who measure them (Alice 1, Alice 2,...), and rules that determine the probabilities observed by the parties given the resources they measure (for us, the decision trees of section 2 and the induced probability rule (9)). Then the theory is *causal* if it satisfies the following conditions: first, given a subset of parties along with all the resources they share—some of which may be additionally shared with parties outside the subset—the subset parties' marginal distribution will be the same regardless of how the measured resources are connected to (or disconnected from, or re-connected to) parties outside the subset: for example, in figure 2, if we look at just the two parties Alice and Bob, their marginal distribution should be the same even if resource $R_1$ is connected to one Charlie-type party while resource $R_2$ is measured by a different Charlie-type party (where the different Charlies, in turn, are perhaps measuring different $R_1$ and $R_2$ resources connected to other Alice and Bob-type parties, etc). A 'Charlie-type' party is a measuring party with the same decision tree.

The second condition, somewhat implicit in the wording of the first given above, is that it makes sense to speak of multiple copies of $R_1$ and $R_2$ resources: the theory should allow the devices to be replicated, and if a network is sound in the sense that each resource is always connected to an appropriate party and vice versa (for example, a resource like $R_2$ in figure 2 is always connected to a party like Alice and a party like Charlie; Charlie is always connected to a resource of form $R_1$ and $R_2$), then the theory provides a sound probability distribution for the parties of the network. (This is key for the mechanics of the inflation technique: a network of interest is 'inflated' to a larger network locally isomorphic to the original one; straightforward constraints on the probability distribution of the larger network reveal subtler insights about the smaller network—for this to work, it is necessary that the larger network *have* a probability distribution.)

The third condition is independence of distributions among parties with no common resources: if two parties measure no common resources, then their joint distribution should factor: $\mathbb{P}(\mathbf{AB}|\mathbf{XY}) = \mathbb{P}(\mathbf{A}|\mathbf{X})\mathbb{P}(\mathbf{B}|\mathbf{Y})$ if there is no resource $R$ measured by both Alice and Bob. This condition is also required to hold more generally for two disjoint subsets of parties.

We now discuss how the theory of wired nonsignaling resources satisfies the conditions of causality described above. First, the nonsignaling condition derived in section 3.1 ensures that the marginal distribution of a subset of parties does not depend on how the resources they share are connected to parties outside the subset. Specifically, looking back to (17), we see that if we remove Alice 1 from consideration, the marginal distribution of the remaining parties is the product of the resource distributions $R(\cdots|\cdots)$ with Alice 1's variables removed. Thus if we have a set of $n' < n$ parties of interest, we can remove the non-$n'$ parties one by one until only the $n'$ parties remain, at which point their marginal distribution is (uniquely) determined as the product of the resources they share, now treated as the marginal resources $R(a_1,\ldots,a_{n'}|x_1,\ldots,x_{n'})$ that remain when the other parties' $a_i$ and $x_i$ variables have been removed. This will be the same expression regardless of how the removed parties were connected or not connected to resources, with the resulting marginal distribution only depending on the decision trees of the $n'$ parties.

The second condition (device replication) follows from the soundness of the probability formula (9) for always producing a consistent probability distribution from cascaded measurements of nonsignaling resources $R_k$, as discussed at length and proved in section 2.3 and appendix A. Finally, the third condition (independence) follows from (9) when we consider that

$$\begin{aligned}\prod_{k=1}^{m} & R_k\left(a_{k_1},\ldots,a_{k_{n_k}}|x_{k_1},\ldots,x_{k_{n_k}}\right)\\ &= \prod_{k\in\mathcal{S}} R_k\left(a_{k_1},\ldots,a_{k_{n_k}}|x_{k_1},\ldots,x_{k_{n_k}}\right)\prod_{k\in\mathcal{S}^C} R_k\left(a_{k_1},\ldots,a_{k_{n_k}}|x_{k_1},\ldots,x_{k_{n_k}}\right)\end{aligned}$$

and if each party either shares resources only from $\mathcal{S}$, or only from $\mathcal{S}^C$, then the factors above will correspond to respective distributions $\mathbb{P}_{\mathcal{S}}$ and $\mathbb{P}_{\mathcal{S}^C}$ that would obtain individually from two disjoint networks treated separately, so that $\mathbb{P}$ factors into the product of $\mathbb{P}_{\mathcal{S}}$ and $\mathbb{P}_{\mathcal{S}^C}$.

## 4. An application: deriving the Chao–Reichardt inequality [18] and others from that of Mao *et al* [19]

We now turn to inequalities witnessing LOSR-GMNL in the three-party setting. Chao and Reichardt [18] give an early example of a constraint on three-party behaviors induced by wired networks of 2-party-only nonlocal nonsignaling resources, with access to global shared (local) randomness; a linear version of this constraint is given in [11] where it is derived rigorously. The arguments in [11, 18] directly work with the nonsignaling resources and do not invoke the inflation technique. Later constraints introduced by [12, 16] and improved upon in [19, 20] employed the inflation technique and so constrain a more general class of behaviors (allowing for additional features in the bipartite-only networks such as entangled measurements of quantum resources).

In the context of the previous section, which solidifies the applicability of the inflation technique to wired nonsignaling boxes, it is notable to show how the early inequality of [11, 18] can be obtained from that of [19]: this exercise provides an alternate proof of the Chao–Reichardt inequality, and shows it constrains a more general class of theories (i.e. all causal theories as opposed to just wired nonsignaling boxes). Let us stipulate that each of the parties have two settings and outcomes, where the settings $X, Y, Z$ take a value in $\{0,1\}$ and the outcomes $A, B, C$

take a value in $\{-1,+1\}$. (Note in previous sections we denoted $A,B,C,X,Y,Z$ with boldface; we do not do so here to align with the notation of [19].) Define

$$\langle A_xB_y\rangle = \mathbb{P}(A=B|xy) - \mathbb{P}(A\neq B|xy)$$

so that the above is equal to the expected value $\mathbb{E}(AB|xy)$ of the product of the outcomes of $A$ and $B$. Similarly, we can define a three-way expectation

$$\langle A_xB_yC_z\rangle = \mathbb{P}(ABC=+1|xyz) - \mathbb{P}(ABC=-1|xyz).$$

Now the inequality of Mao *et al* [19] is

$$\langle A_0B_0\rangle + \langle A_0B_1\rangle + \langle A_1B_0C_1\rangle - \langle A_1B_1C_1\rangle + 2\langle A_0C_0\rangle \leqslant 4 \tag{21}$$

(see expression (3) in this reference), where the above must hold of networks of bipartite-only nonlocal resources but can be violated if the three-way entangled GHZ resource is measured. The 3-party inequality of Chao and Reichardt [18] is formulated in [11] (see expressions (14) and (15) therein) as

$$\begin{aligned}&4\mathbb{P}(A\neq C|X=0,Z=0) + P(A\neq B|X=0,Y=0) + P(A\neq B|X=0,Y=1)\\&\quad + \mathbb{P}(ABC=-1|X=1,Y=0,Z=1) + \mathbb{P}(ABC=+1|X=1,Y=1,Z=1) \geqslant 1\end{aligned} \tag{22}$$

which if we re-write in terms of $\langle AB\rangle$ type expressions using conversions of the form $\mathbb{P}(A=B|xy) = (1+\langle A_xB_y\rangle)/2$ and $\mathbb{P}(A\neq B|xy) = (1-\langle A_xB_y\rangle)/2$ along with their three-party analogs, and perform some algebra, we get

$$\langle A_0B_0\rangle + \langle A_0B_1\rangle + \langle A_1B_0C_1\rangle - \langle A_1B_1C_1\rangle + 4\langle A_0C_0\rangle \leqslant 6$$

which can be obtained from (21) by adding the trivial algebraic inequality $\langle A_0C_0\rangle \leqslant 1$ twice. Equation (21) is evidently the stronger constraint.

Interestingly, the inequality of Cao *et al* [20] can be derived from (21) as well: If we re-label Bob's outcomes when his setting is 1 by interchanging $+1$ and $-1$, (21) becomes

$$\langle A_0B_0\rangle - \langle A_0B_1\rangle + \langle A_1B_0C_1\rangle + \langle A_1B_1C_1\rangle + 2\langle A_0C_0\rangle \leqslant 4. \tag{23}$$

Switching the roles of Alice and Charlie in (23) yields

$$\langle C_0B_0\rangle - \langle C_0B_1\rangle + \langle A_1B_0C_1\rangle + \langle A_1B_1C_1\rangle + 2\langle A_0C_0\rangle \leqslant 4, \tag{24}$$

and adding (23) and (24) together yields

$$\langle A_0B_0\rangle + \langle B_0C_0\rangle - \langle A_0B_1\rangle - \langle B_1C_0\rangle + 4\langle A_0C_0\rangle + 2\langle A_1B_0C_1\rangle + 2\langle A_1B_1C_1\rangle \leqslant 8 \tag{25}$$

which is the 3-party inequality (1) in [20]. Thus all known $(3,2,2)$ inequalities (3-party, 2-outcome, 2-setting) can be derived from that of Mao *et al* [19]. Note that since both (23) and (24) require genuine tripartite nonlocality to violate, their sum (25) should not necessarily be considered a weaker witness of LOSR-GMNL when compared to (21).

Cao *et al* [20] contains another inequality (S14) in the supplementary material which in the 3-party case is not a $(3,2,2)$ inequality (Bob has a third setting) but it can be obtained from (21) nonetheless. The three party version of (S14) is

$$\frac{1-\langle C_1\rangle}{2}\left(\langle A_0B_0\rangle_{C=-1,Z=1}-\langle A_0B_1\rangle_{C=-1,Z=1}+\langle A_1B_0\rangle_{C=-1,Z=1}+\langle A_1B_1\rangle_{C=-1,Z=1}\right)$$
$$+\frac{1+\langle C_1\rangle}{2}\left(\langle A_0B_0\rangle_{C=+1,Z=1}+\langle A_0B_1\rangle_{C=+1,Z=1}+\langle A_1B_0\rangle_{C=+1,Z=1}-\langle A_1B_1\rangle_{C=+1,Z=1}\right)$$
$$+\langle A_0B_2\rangle+\langle B_2C_0\rangle\leqslant 6, \tag{26}$$

where $\langle A_xB_y\rangle_{C=c,Z=1}$ is the expectation conditioned on $C=c, Z=1$. It turns out that the first two lines of (26) are equivalent to the first four terms of (21), and the fifth term of (21) can be replaced using the algebraic inequality $\langle A_0C_0\rangle\geqslant\langle A_0B_2\rangle+\langle B_2C_0\rangle-1$ (this inequality is used in [16, 20, 25] and can be confirmed by writing out all the probabilities), leading to (26). Details of the derivation are given in appendix C.

The only other three party inequality currently known to witness LOSR-GMNL is (1) in [16], which was tested in [21]. Like (26), this inequality has a third setting for Bob, and while (1) of [16] admits a linear form [31] it does not appear to be directly derivable from (21).

## 5. Conclusion

We have shown the consistency of probability distributions induced by wired nonsignaling resources, shown that such distributions are themselves nonsignaling, and discussed other properties such as causality and the ability to factor out classical random resources and ignore local deterministic distributions while restricting attention to extremal nonsignaling resources. This study was motivated in part by new definitions of Genuine Multipartite Nonlocality (the 'LOSR-GMNL' definition of [12]), and we closed with an example showing how most inequalities witnessing tripartite GMNL can be derived from that of [19]. Going forward, the framework developed in this paper will provide a useful foundation for rigorously proving future results about wired nonsignaling resources; this will be useful in studying the gap between this scenario and more general scenarios permitting entangled measurements—notably, the inflation technique constrains all causal theories and so cannot directly target this gap. The results here are also relevant to other paradigms, such as for example networks of quantum-achievable nonsignaling resources measured in cascaded fashion as studied in the proposed definition of genuine network nonlocality given in [28]. Future work may also explore generalizations to encompass resources that admit some restricted form of signaling, such as in models that utilize underlying one-way signaling resources to replicate quantum nonlocal behaviors [15, 32], to see under what weaker conditions a consistent joint distribution as in (9) may still be guaranteed.

## Data availability statement

No new data were created or analyzed in this study.

## Acknowledgments

This work was partially supported by NSF Award No. 2210399 and AFOSR Award No. FA9550-20-1-0067.

## Appendix A. Proofs of nonsignaling properties

Here we write out a couple of the longer equation sequences proving claims in section 2.1. First, we write out how (3) is a consequence of (2). The three party version of this argument appears in section IIIA of [10] which we merely iterate more times to obtain the following:

$$
\begin{aligned}
&\sum_{\vec{a}_q} R\left(\vec{a}_p, \vec{a}_q | \vec{x}_p, \vec{x}_q\right) \\
&\quad = \sum_{a_{p+1},\dots,a_n} R\left(\vec{a}_p, a_{p+1}, \dots, a_n | \vec{x}_p, x_{p+1}, \dots, x_n\right) \\
&\quad = \sum_{a_{p+1},\dots,a_{n-1}} \left( \sum_{a_n} R\left(\vec{a}_p, a_{p+1}, \dots, a_{n-1}, a_n | \vec{x}_p, x_{p+1}, x_{p+2}, \dots, x_{n-1}, x_n\right) \right) \\
&\quad = \sum_{a_{p+1},\dots,a_{n-1}} \left( \sum_{a_n} R\left(\vec{a}_p, a_{p+1}, \dots, a_{n-1}, a_n | \vec{x}_p, x_{p+1}, x_{p+2}, \dots, x_{n-1}, x_n'\right) \right) \\
&\quad = \sum_{a_{p+1},\dots,a_{n-2},a_n} \left( \sum_{a_{n-1}} R\left(\vec{a}_p, a_{p+1}, \dots, a_{n-1}, a_n | \vec{x}_p, x_{p+1}, x_{p+2}, \dots, x_{n-1}, x_n'\right) \right) \\
&\quad = \sum_{a_{p+1},\dots,a_{n-2},a_n} \left( \sum_{a_{n-1}} R\left(\vec{a}_p, a_{p+1}, \dots, a_{n-1}, a_n | \vec{x}_p, x_{p+1}, x_{p+2}, \dots, x_{n-1}', x_n'\right) \right) \\
&\quad\vdots \\
&\quad = \sum_{a_{p+2},\dots,a_n} \left( \sum_{a_{p+1}} R\left(\vec{a}_p, a_{p+1}, \dots, a_{n-1}, a_n | \vec{x}_p, x_{p+1}, x_{p+2}', \dots, x_{n-1}', x_n'\right) \right) \\
&\quad = \sum_{a_{p+2},\dots,a_n} \left( \sum_{a_{p+1}} R\left(\vec{a}_p, a_{p+1}, \dots, a_{n-1}, a_n | \vec{x}_p, x_{p+1}', x_{p+2}', \dots, x_{n-1}', x_n'\right) \right) \\
&\quad = \sum_{\vec{a}_q} R\left(\vec{a}_p, \vec{a}_q | \vec{x}_p, \vec{x}_q'\right).
\end{aligned}
$$

The steps above alternate between re-arranging order of summation, and then applying (2) to the inner sum within parentheses. The above proof does not depend on the ordering of the parties; choosing $\vec{x}_p$ as an initial string just makes it easier to notate. The condition thus applies to any two complementary sets of parties.

We now show that our definition of $R(\vec{a}_p|\vec{x}_p)$ in (5) is consistent with what we would find for this quantity from a direct manipulating conditional probabilities, if we model the inputs $X_i$ as random variables with a probability distribution (which along with the specification of $R(\vec{A}_n|\vec{X}_n)$ induces a full joint distribution of $\vec{A}_n$ and $\vec{X}_n$). Assuming $R(\vec{x}_p) > 0$—if not, $R(\vec{a}_p|\vec{x}_p)$ is undefined—we can write, for any choice $\vec{a}_p$,

$$
\begin{aligned}
R\left(\vec{a}_p|\vec{x}_p\right) &= R\left(\vec{a}_p,\vec{x}_p\right)/R\left(\vec{x}_p\right) \\
&= \left[\sum_{\vec{x}_q:R(\vec{x}_p,\vec{x}_q)>0} R\left(\vec{a}_p,\vec{x}_p,\vec{x}_q\right)\right]/R\left(\vec{x}_p\right) \\
&= \left[\sum_{\vec{x}_q:R(\vec{x}_p,\vec{x}_q)>0} R\left(\vec{a}_p|\vec{x}_p,\vec{x}_q\right)R\left(\vec{x}_p,\vec{x}_q\right)\right]/R\left(\vec{x}_p\right) \\
&= \left[\sum_{\vec{x}_q:R(\vec{x}_p,\vec{x}_q)>0} R\left(\vec{a}_p|\vec{x}_p,\vec{x}_q^*\right)R\left(\vec{x}_p,\vec{x}_q\right)\right]/R\left(\vec{x}_p\right) \\
&= \left[R\left(\vec{a}_p|\vec{x}_p,\vec{x}_q^*\right)\sum_{\vec{x}_q:R(\vec{x}_p,\vec{x}_q)>0} R\left(\vec{x}_p,\vec{x}_q\right)\right]/R\left(\vec{x}_p\right) \\
&= \left[R\left(\vec{a}_p|\vec{x}_p,\vec{x}_q^*\right)R\left(\vec{x}_p\right)\right]/R\left(\vec{x}_p\right) \\
&= R\left(\vec{a}_p|\vec{x}_p,\vec{x}_q^*\right),
\end{aligned}
$$

where $\vec{x}_q^*$ is a fixed choice of values for $\vec{X}_q$, which allows for pulling the term $R(\vec{a}_p|\vec{x}_p,\vec{x}_q^*)$ out of the sum over $\vec{x}_q$ after previously replacing each (varying) choice of $\vec{x}_q$ in $R(\vec{a}_p|\vec{x}_p,\vec{x}_q)$ with this fixed $\vec{x}_q^*$ by invoking (4). Since $\vec{x}_q^*$ above can be any value of $\vec{x}_q$ for which $R(\vec{x}_p,\vec{x}_q) > 0$, defining $R(\vec{a}_p|\vec{x}_p)$ in (5) as $R(\vec{a}_p|\vec{x}_p,\vec{x}_q)$ for any fixed choice of $\vec{x}_q$ is sensible.

## Appendix B. Normalization in the general setting

To prove normalization of (9) in the general case, we rely on the following arithmetic construction. Suppose a quantity $Q$ can be written as

$$Q = \sum_i \xi_i f(i), \quad \text{with} \sum_i \xi_i = 1 \tag{B.1}$$

where each $f(i)$ is a number which may vary with $i$. If $f(i)$ happens to equal 1 for all choices of $i$, then $Q = 1$, but we do not initially assume this is the case. We assume instead that each $f(i)$ can be written in a form parallel to (B.1):

$$f(i) = \sum_j \eta_j^i g^i(j), \quad \text{with} \sum_j \eta_j^i = 1$$

where $g^i(j)$ is a number that may vary with $j$, and then we say that $Q$ satisfies the *recursive sum-to-1 property* if the process can always be repeated such that each new nested functional term can be written in the form of (B.1), while assuming that this process eventually terminates in a final expression of the form (B.1) where the functional term *does* equal one uniformly (i.e. not varying with the summed index). Then with a little thought, we see that the original quantity $Q$ must equal one as follows: each bottom-level sum, for which the functional term is uniformly one, will itself equal one; then move back up one level where the functional terms are now known to be 1, and the next-higher-level sum will equal one as well; continuing to recursively work back up to the original quantity $Q$ level by level, we find $Q = 1$.

For our problem, we show that for a fixed setting choice $\mathbf{x}_1,\ldots,\mathbf{x}_n$, the sum of all $\mathbb{P}(\vec{a}_1,\ldots,\vec{a}_m|\mathbf{x}_1,\ldots,\mathbf{x}_n)$ terms, which is the quantity

$$\sum_{\vec{a}_1,\ldots,\vec{a}_m}\prod_{k=1}^{m}R_k\left(a_{k_1}^{(k)},\ldots,a_{k_{n_k}}^{(k)}|x_{k_1}^{(k)},\ldots,x_{k_{n_k}}^{(k)}\right), \tag{B.2}$$

satisfies this recursive sum-to-1 property, and thus equals 1. The idea is to successively perform the manipulation that took us from (12) to (14) as we work down a decision tree until eventually what is left in the inner sum is a sum over a single variable that is equal to one.

As a first step, pick a party $p$—for ease of notation, let us say it is Alice 1—and consult this party's decision tree to find the first resource they consult after being provided setting $\mathbf{x}_1$; denote this $R_t$ and let $x_{t_1}^{(t)}=x_1^{(t)}$ denote the input specified by $inp_1$. Now in (B.2), limit the sum over $a_1^{(t)}$ to precisely those values for which $R_t(a_1|x_1)>0$, which does not change the value of the sum as $R_t(a_1|x_1)=0$ implies that the term $R_t(a_{t_1},a_{t_2},\ldots,a_{t_{n_t}}|x_{t_1},x_{t_2},\ldots,x_{t_{n_t}})$ appearing in the summand will be zero as well. (Recall that $t_1,t_2,\ldots,t_{n_t}$ denotes the indices of the subset of parties sharing resource $R_t$, so for this resource $a_{t_1}^{(t)}=a_1^{(t)}$ and $x_{t_1}^{(t)}=x_1^{(t)}$.) Now for each value of $a_1^{(t)}$ for which $R_t(a_1|x_1)>0$ we can write

$$R_t\left(a_1,a_{t_2},\ldots,a_{t_{n_t}}|x_1,x_{t_2},\ldots,x_{t_{n_t}}\right)=R_t\left(a_{t_2},\ldots,a_{t_{n_t}}|x_1,x_{t_2},\ldots,x_{t_{n_t}},a_1\right)R_t\left(a_1|x_1\right)$$

via the same manipulation that was performed in (13). Then since $x_1$ is determined by $\mathbf{x}_1$ alone, we can pull $R_t(a_1|x_1)$ out of the sum and re-write (B.2) as follows:

$$\underbrace{\sum_{a_1^{(t)}:R_t(a_1|x_1)>0}R_t\left(a_1|x_1\right)}_{\sum_i\xi_i}\underbrace{\sum_{\{\vec{a}_1,\ldots,\vec{a}_m\}\setminus a_1^{(t)}}R_1\left(\cdots|\cdots\right)R_2\left(\cdots|\cdots\right)\cdots R_k\left(\cdots|\cdots\right)}_{f(i)}, \tag{B.3}$$

where the $R(\cdots|\cdots)$ terms of the inner sum are as in (B.2) except for $R_t(\cdots|\cdots)$ which now equals $R_t(a_{t_2},\ldots,a_{t_{n_t}}|x_1,x_{t_2},\ldots,x_{t_{n_t}},a_1)$. Now, following (5) the terms $R_t(a_1|x_1)$ constitute a probability distribution and so will sum to one, justifying the labeling with $\sum_i\xi_i$ above, so the above expression is a $Q$-type quantity as in (B.1). The value of the inner sum can vary with the choice of index of the outer sum, as is allowed for the $f(i)$ terms in (B.1).

To perform the inductive step of the argument, we show that that the terms labeled $f(i)$ in (B.3) satisfy certain general conditions, and that these conditions (alone) ensure that each $f(i)$ term can always be re-written in a form $\sum_j\eta_j g(j)$ with $\sum_j\eta_j=1$ *such that* that the same general conditions will hold for each $g(j)$; thus the process will always be repeatable, and as a final step we will see that it terminates in an expression uniformly equaling one. The general conditions are motivated by the idea that we will pull out terms from the inner sum one by one, with each pull-out corresponding to taking a single step down a party's decision tree to the next consulted resource.

Now we lay out the conditions: each term labeled $f(i)$ in (B.3) (which vary with the outer sum) is an expression of the form

$$\sum_{M\subseteq\{\vec{a}_1,\ldots,\vec{a}_m\}}R_1\left(\cdots|\cdots\right)\cdots R_m\left(\cdots|\cdots\right) \tag{B.4}$$

where summing over $M\subseteq\{\vec{a}_1,\ldots,\vec{a}_m\}$ is to be understood that that a subcollection of variables of the form $a_j^{(k)}$ are being summed over. If we think of ourselves as working down decision trees, $M$ corresponds to parties' 'pending outputs' from resources that have not yet been consulted. The following properties are satisfied by expression (B.4):

(i) For each $a_p^{(k)} \in M^C$—that is, an $a_p^{(k)}$ that is not being summed over—a fixed choice $a_p^{(k)}$ appears in the conditioner of $R_k(\cdots|\cdots)$, along with a fixed choice of $x_p^{(k)}$, and these do not vary. (These correspond to resources that have already been consulted, having worked part way down a path on a decision tree.)

(ii) For each party $p$, if we collect all the fixed values $a_p^{(k)}$ from $M^C$ for this choice of $p$, these determine an initial path in party $p$'s decision tree descending from the overall setting choice $\mathbf{x}_p$. The fixed $x_p^{(k)}$ appearing inside the summand are consistent with the $inp_i$ on this initial path.

(iii) For each $a_j^{(k)}$ in $M$, $a_j^{(k)}$ appears (varying) in the front of the appropriate $R_k(\cdots|\cdots)$ term, and the corresponding $x_j^{(k)}$ will appear in the conditioner. These $x_j^{(k)}$ are not necessarily fixed and may change as the sum over $M$ is performed—they are determined by the fixed choice of $a$ values from $M^C$ along with the varying-with-the-sum choices of $a$ values from $M$.

(iv) We adopt a convention that any term of the form $R_k(\emptyset|\cdots)$—that is, with no terms in the front—equals one. This corresponds to a resource that all parties have already consulted and so the corresponding $a_j^{(k)}$ are all in $M^C$. (Operationally, this should be understood to indicate a resource $R_k$ that has been pulled out of the inner sum completely; however we leave a rump term behind with this notational oddity to help maintain the inductive form (B.4) through all steps.)

Now we show that the conditions ensure that (B.4) can be re-written as $\sum_j \eta_j g(j)$ with $\sum_j \eta_j = 1$ and each $g(j)$ also satisfying the conditions. To do so, consult the part of a party $p$'s decision tree that is determined by that party's initial setting $\mathbf{x}_p$ along with the fixed choices of that party's $a$ values from the collection $M^C$ (if any), which by conditions (i) and (ii) does determine a unique (initial) path for party $p$. Let $i$ be the length of this initial path. Then it will determine a choice of resource $c_i$ and input $inp_i$ to be used at the next step; thus for the resource $R_t(\cdots|\cdots)$ with $t = c_i$, the value of $x_p^{(t)}$ in the conditioner will be fixed as $inp_i$ for all terms of the sum in (B.4) (even as the corresponding $a_p^{(t)} \in M$ varies as it is summed over). For ease of notation let us assume that $p = 1$, so $R_t(\cdots|\cdots)$ will appear in (B.4) as

$$R_t(a_1, \vec{a}_q | x_1, \vec{x}_q, \vec{x}_r, \vec{a}_r) = R_t^{\vec{x}_r, \vec{a}_r}(a_1, \vec{a}_q | x_1, \vec{x}_q) \tag{B.5}$$

where $\vec{a}_q$ are among the $M$ indices and $\vec{a}_r$ are among the $M^C$ indices. For values of $a_1$ for which $R_t(a_1|x_1, \vec{x}_r, \vec{a}_r) = R_t^{\vec{x}_r, \vec{a}_r}(a_1|x_1)$ is nonzero, we can apply (7) to $R_t^{\vec{x}_r, \vec{a}_r}$ to re-write the above expression as

$$R_t(a_1, \vec{a}_q | x_1, \vec{x}_q, \vec{x}_r, \vec{a}_r) = R_t(\vec{a}_q | x_1, \vec{x}_q, \vec{x}_r, a_1, \vec{a}_r) R_t(a_1 | x_1, \vec{x}_r, \vec{a}_r)$$

where the equality follows from the fact that this sort of conditioning can be performed iteratively as discussed following (8). Now pull out $R_t(a_1|x_1, \vec{x}_r, \vec{a}_r)$ to re-write (B.4) as follows:

$$\sum_{a_1^{(t)}: R_t(a_1|x_1, \vec{x}_r, \vec{a}_r) > 0} R_t(a_1|x_1, \vec{x}_r, \vec{a}_r) \sum_{M' = M \setminus a_1^{(t)}} R_1(\cdots|\cdots) \cdots R_m(\cdots|\cdots) \tag{B.6}$$

where the inner summand above differs from the summand in (B.4) by replacing $R_t(a_1, \vec{a}_q | x_1, \vec{x}_q, \vec{x}_r, \vec{a}_r)$ with $R_t(\vec{a}_q | x_1, \vec{x}_q, \vec{x}_r, a_1, \vec{a}_r)$. Now, the first sum in (B.6) is the $\sum_i \xi_i$ portion of (B.1), where $\sum_i \xi_i = 1$ as $R_t(a_1 | x_1, \vec{x}_r, \vec{a}_r)$ is a probability distribution over $a_1$. And we now can argue that for each fixed choice of $a_1^{(t)}$,

$$\sum_{M' = M \setminus a_1^{(t)}} R_1(\cdots|\cdots) \cdots R_m(\cdots|\cdots) \tag{B.7}$$

is an expression of the form (B.1) satisfying the general conditions (i)–(iv), where we have effectively moved $a_1^{(t)}$ from $M$ to $M^C$. To elaborate: as required by (i) for $M'^C$, a fixed choice of $a_1^{(t)}$ now appears in the *conditioner* of $R_t(\cdots|\cdots)$, and as noted preceding (B.5) the choice of $x_1^{(k)}$ will be fixed as well. The initial path for Alice 1 referred to in (ii) is now one level longer in (B.7) compared to (B.4), while still satisfying the condition. For all other parties, satisfaction of (i) and (ii) in (B.4) carries over immediately to (B.7). Finally, satisfaction of (iii) carries over directly from (B.4) to (B.7) as $M' \subset M$.

Regarding the eventual termination of this process, each round of induction moves an $a$ variable from the front of an $R_k(\cdots|\cdots)$ term to the conditioner; there is a finite number of times this will occur before all terms remaining in (B.7) are of the form described in condition (iv)—and so $M'$ is the empty set—at which point (B.7) equals 1, completing the argument.

## Appendix C. Obtaining (S14) in [20] from (1) in [19]

In this appendix we explain how the first two lines of (26) are equivalent to the first four terms of (21). Thus the expression (26) is a consequence of (21) when $2\langle A_0C_0\rangle$ is replaced according to an algebraic inequality described in the main text.

Substituting $(1+\langle C_1\rangle)/2 = \mathbb{P}(C=+1|Z=1)$ and $(1-\langle C_1\rangle)/2 = \mathbb{P}(C=-1|Z=1)$, we rewrite (26) as

$$\begin{aligned}
&\mathbb{P}(C=-1|Z=1)\left(\langle A_0B_0\rangle_{C=-1,Z=1} + \langle A_0B_1\rangle_{C=-1,Z=1} - \langle A_1B_0\rangle_{C=-1,Z=1} + \langle A_1B_1\rangle_{C=-1,Z=1}\right)\\
&\quad + \mathbb{P}(C=+1|Z=1)\left(\langle A_0B_0\rangle_{C=+1,Z=1} + \langle A_0B_1\rangle_{C=+1,Z=1} + \langle A_1B_0\rangle_{C=+1,Z=1} - \langle A_1B_1\rangle_{C=+1,Z=1}\right)\\
&\quad + \langle A_0B_2\rangle + \langle B_2C_0\rangle \leqslant 6.
\end{aligned} \tag{C.1}$$

Writing out expectations in a more explicit form

$$\langle A_xB_y\rangle_{C=c,Z=1} = \mathbb{E}(AB|X=x,Y=y,Z=1;C=c)$$

and noting that by no-signaling

$$\mathbb{P}(C=c|Z=1) = \mathbb{P}(C=c|X=x,Y=y,Z=1),$$

we can rewrite the sum of two terms $\mathbb{P}(C=-1|Z=1)\langle A_0B_0\rangle_{C=-1,Z=1}$ and $\mathbb{P}(C=+1|Z=1)\langle A_0B_0\rangle_{C=+1,Z=1}$ that appear in (C.1) after multiplying out as

$$\begin{aligned}
&\mathbb{E}(AB|X=0,Y=0,Z=1;C=-1)\,\mathbb{P}(C=-1|X=0,Y=0,Z=1)\\
&\quad + \mathbb{E}(AB|X=0,Y=0,Z=1;C=+1)\,\mathbb{P}(C=+1|X=0,Y=0,Z=1)\\
&\quad = \mathbb{E}(AB|X=0,Y=0,Z=1)
\end{aligned}$$

by the law of iterated expectation; $\mathbb{E}(AB|X=0,Y=0,Z=1)$ is equal to $\mathbb{E}(AB|X=0,Y=0) = \langle A_0B_0\rangle$ in turn by nonsignaling. A similar argument yields $\langle A_0B_1\rangle$ from two other terms of (C.1) after multiplying out. On the other hand, we can write $-\langle A_1B_0\rangle_{C=-1,Z=1}$ as

$$-\left[\mathbb{P}(A=B|X=1,Y=0,Z=1;C=-1) - \mathbb{P}(A\neq B|X=1,Y=0,Z=1;C=-1)\right]$$

which when we multiply by $\mathbb{P}(C=-1|Z=1) = \mathbb{P}(C=-1|X=1,Y=0,Z=1)$ becomes

$$-\mathbb{P}(A=B,C=-1|X=1,Y=0,Z=1) + \mathbb{P}(A\neq B,C=-1|X=1,Y=0,Z=1);$$

performing a similar calculation on $\langle A_1B_0\rangle_{C=+1,Z=1}\mathbb{P}(C=+1|Z=1)$ and adding the result to the expression above, the resulting sum can be equivalently re-written as

$$\mathbb{P}(ABC=+1|X=1,Y=0,Z=1) - \mathbb{P}(ABC=-1|X=1,Y=0,Z=1) = \langle A_1B_0C_1\rangle.$$

Similarly, the final remaining two $\langle A_xB_y\rangle$ type terms in (C.1) are equivalent to $-\langle A_1B_1C_1\rangle$.

## ORCID iD

Peter Bierhorst https://orcid.org/0000-0003-2781-5448

## References

[1] Bell J 1964 On the Einstein Podolsky Rosen paradox *Physics* **1** 195–200
[2] Clauser J, Horne A, Shimony A and Holt R 1969 Proposed experiment to test local hidden-variable theories *Phys. Rev. Lett.* **23** 880–4
[3] Hensen B *et al* 2015 Loophole-free Bell inequality violation using electron spins separated by 1.3 km *Nature* **526** 682
[4] Shalm L K *et al* 2015 Strong loophole-free test of local realism *Phys. Rev. Lett.* **115** 250402
[5] Giustina M *et al* 2015 Significant-loophole-free test of Bell's theorem with entangled photons *Phys. Rev. Lett.* **115** 250401
[6] Rosenfeld W, Burchardt D, Garthoff R, Redeker K, Ortegel N, Rau M and Weinfurter H 2017 Event-ready Bell test using entangled atoms simultaneously closing detection and locality loopholes *Phys. Rev. Lett.* **119** 010402
[7] Barrett J and Pironio S 2005 Popescu-Rohrlich correlations as a unit of nonlocality *Phys. Rev. Lett.* **95** 140401
[8] Popescu S and Rohrlich D 1994 Quantum nonlocality as an axiom *Found. Phys.* **24** 379–85
[9] Bierhorst P *et al* 2018 Experimentally generated randomness certified by the impossibility of superluminal signals *Nature* **557** 223–6
[10] Barrett J, Linden N, Massar S, Pironio S, Popescu S and Roberts D 2005 Nonlocal correlations as an information-theoretic resource *Phys. Rev.* A **71** 022101
[11] Bierhorst P 2021 Ruling out bipartite nonsignaling nonlocal models for tripartite correlations *Phys. Rev.* A **104** 012210
[12] Coiteux-Roy X, Wolfe E and Renou M-O 2021 No bipartite-nonlocal causal theory can explain nature's correlations *Phys. Rev. Lett.* **127** 200401
[13] Svetlichny G 1987 Distinguishing three-body from two-body nonseparability by a Bell-type inequality *Phys. Rev.* D **35** 3066
[14] Seevinck M and Svetlichny G 2002 Bell-type inequalities for partial separability in N-particle systems and quantum mechanical violations *Phys. Rev. Lett.* **89** 060401
[15] Bancal J-D, Barrett J, Gisin N and Pironio S 2013 Definitions of multipartite nonlocality *Phys. Rev.* A **88** 014102
[16] Coiteux-Roy X, Wolfe E and Renou M-O 2021 Any physical theory of nature must be boundlessly multipartite nonlocal *Phys. Rev.* A **104** 052207
[17] Greenberger D, Horne M and Zeilinger A 1989 Going beyond Bell's theorem *Bell's Theorem, Quantum Theory and Conceptions of the Universe* ed M Kaftos (Kluwer Academic) pp 69–72

[18] Chao R and Reichardt B 2017 Test to separate quantum theory from non-signaling theories (arXiv:1706.02008 [quant-ph])
[19] Mao Y-L, Li Z-D, Yu S and Fan J 2022 Test of genuine multipartite nonlocality *Phys. Rev. Lett.* **129** 150401
[20] Cao H, Renou M-O, Zhang C, Massé G, Coiteux-Roy X, Liu B-H, Huang Y-F, Li C-F, Guo G-C and Wolfe E 2022 Experimental demonstration that no tripartite-nonlocal causal theory explains nature's correlations *Phys. Rev. Lett.* **129** 150402
[21] Huang L *et al* 2022 Experimental demonstration of genuine tripartite nonlocality under strict locality conditions *Phys. Rev. Lett.* **129** 060401
[22] Janotta P 2012 Generalizations of boxworld *Proc. Theor. Comput. Sci.* **95** 183
[23] Chiribella G, D'Ariano G M and Perinotti P 2011 Informational derivation of quantum theory *Phys. Rev.* A **84** 012311
[24] Chiribella G 2014 Dilation of states and processes in operation-probabilistic theories *Electron. Proc. Theor. Comput. Sci.* **172** 1
[25] Wolfe E, Spekkens R and Fritz T 2019 The inflation technique for causal inference with latent variables *J. Causal Inference* **7** 2
[26] Bierhorst P and Prakash J 2023 Hierarchy of multipartite nonlocality and device-independent effect witnesses *Phys. Rev. Lett.* **130** 250201
[27] Rabelo R, Ho M, Cavalcanti D, Brunner N and Scarani V 2011 Device-independent certification of entangled measurements *Phys. Rev. Lett.* **107** 050502
[28] Šupić I, Bancal J-D, Cai Y and Brunner N 2022 Genuine network quantum nonlocality and self-testing *Phys. Rev.* A **105** 022206
[29] Gisin N, Bancal J-D, Cai Y, Remy P, Tavakoli A, Cruzeiro E Z, Popescu S and Brunner N 2020 Constraints on nonlocality in networks from no-signaling and independence *Nat. Commun.* **11** 2378
[30] Pironio S, Bancal J-D and Scarani V 2011 Extremal correlations of the tripartite no-signaling polytope *J. Phys. A: Math. Theor.* **44** 065303
[31] Patra S and Bierhorst P 2024 Strength of statistical evidence for genuine tripartite nonlocality (arXiv:2407.19587 [quant-ph])
[32] Bancal J-D, Pironio S, Acín A, Liang Y-C, Scarani V and Gisin N 2012 Quantum non-locality based on finite-speed causal influences leads to superluminal signalling *Nat. Phys.* **8** 867–70