

This is a repository copy of *Multisystem measurements in generalized probabilistic theories and their role in information processing*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/206686/>

Version: Published Version

Article:

Eftaxias, Giorgos, Weilenmann, Mirjam and Colbeck, Roger orcid.org/0000-0003-3591-0576 (2023) Multisystem measurements in generalized probabilistic theories and their role in information processing. *Physical Review A*. 062212. ISSN 1094-1622

<https://doi.org/10.1103/PhysRevA.108.062212>

Reuse




This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Multisystem measurements in generalized probabilistic theories and their role in information processing

Giorgos Eftaxias ^{1,2,*}, Mirjam Weilenmann ^{3,4,†} and Roger Colbeck ^{2,‡}

¹Quantum Engineering Centre for Doctoral Training, University of Bristol, Bristol BS8 1FD, United Kingdom

²Department of Mathematics, University of York, York YO10 5DD, United Kingdom

³Institute for Quantum Optics and Quantum Information (IQOQI), Austrian Academy of Sciences, Boltzmannngasse 3, 1090 Vienna, Austria

⁴Département de Physique Appliquée, Université de Genève, 1205 Genève, Switzerland



(Received 25 August 2023; accepted 2 November 2023; published 15 December 2023)

Generalized probabilistic theories (GPTs) provide a framework in which a range of possible theories can be examined, including classical theory, quantum theory, and those beyond. In general, enlarging the state space of a GPT leads to fewer possible measurements because the additional states give stronger constraints on the set of effects, the constituents of measurements. This can have implications for information processing. In box world, for example, a GPT in which any no-signaling distribution can be realized, there is no analog of a measurement in the Bell basis and hence the analog of entanglement swapping is impossible. A comprehensive study of measurements on multiple systems in box world has been lacking. Here we consider such measurements in detail, distinguishing those that can be performed by interacting with individual systems sequentially (termed wirings), and the more interesting set of those that cannot. We compute all the possible box-world effects for cases with small numbers of inputs, outputs, and parties, identifying those that are wirings. The large state space of box world leads to a small effect space and hence the effects of box world are widely applicable in GPTs. We also show some possible uses of nonwirings for information processing by studying state discrimination, nonlocality distillation, and the box-world analog of nonlocality without entanglement. Finally, we connect our results to the study of logically consistent classical processes and to the composition of contextuality scenarios. By enhancing understanding of measurements in box world, our results could be useful in studies of possible underlying principles on which quantum theory can be based.

DOI: [10.1103/PhysRevA.108.062212](https://doi.org/10.1103/PhysRevA.108.062212)

I. INTRODUCTION

Standard textbook presentations of the postulates of quantum mechanics usually begin with a list of mathematical axioms with relatively little accompanying explanation. This is in contrast to relativity theory, for example, which can be based on the premise that the laws of physics are frame independent. Whether or how we can formulate quantum theory in a similar way remains open in spite of significant investigation (see, e.g., [1–6]). Quantum theory has some counterintuitive features such as the presence of nonlocal correlations that seemingly defy classical explanation [7]. A general framework to study these features in the context of quantum theory and possible alternatives is that of generalized probabilistic theories (GPTs) [8]. Beyond classical and quantum theory one well-studied GPT is box world, which

allows arbitrary no-signaling distributions to be realized. It is known for instance that particular cryptographic tasks remain possible even against adversaries that have access to box-world systems [9–12]. In a further line of work that tries to single out quantum correlations within a range of alternative theories (see, e.g., [13–16]), box world is a useful foil theory. In spite of these results, the structure of multisystem measurements in box world has not been developed in detail, but understanding these is important to fully characterize the information processing power of box world. Furthermore, although we refer to effects in box world throughout this paper, the measurements we find are applicable to a range of GPTs (see Sec. VIII E).

Further motivation for the study of multisystem measurements comes from the recent trend for studying information processing in quantum networks [17–21]. It is likely that simple quantum networks will be built in the near future, hence, it is useful to explore the possibilities these networks bring. GPTs provide a useful means to form a more general understanding of this. To compare what is possible in quantum theory as opposed to other GPTs it is furthermore necessary to understand the structure of multisystem measurements in the latter. Such comparison furthermore informs the question of in what sense quantum theory is optimal for information processing [22]. In this work, we thus explore such multisystem measurements. In addition, we connect our results to the

*giorgos.eftaxias@bristol.ac.uk

†mirjam.weilenmann@unige.ch

‡roger.colbeck@york.ac.uk

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

study of logically consistent classical processes [23] and of composition of contextuality scenarios [24,25], pointing to a new kind of composition in the latter case.

Our study of the set of possible measurements in box world proceeds using the set of effects, the constituent parts of measurements (see later). The large set of possible states in box world comes at the expense of having a smaller set of effects than in theories with weaker correlations. In this work we will be interested in the set of possible measurements that can be performed with access to several systems in box world. One type of such measurements are the *wirings* [8], which correspond to processes in which a measurement is applied to one system, then to a second depending on the result, and so on (or convex combinations of such processes). Wirings are hence implementable using local operations and classical communication (LOCC) when the individual systems are separated.¹ It is well known that in quantum theory not all measurements are of this form. For example, a measurement in the Bell basis as used in teleportation cannot be realized by wiring together measurements on individual systems. For single and bipartite systems in box world, all measurements are wirings [8], while this is no longer true for three or more systems [26]. One of the aims of this paper is to find the multi-system measurements that are not wirings and investigate their significance for information processing.

We classify scenarios by the number of systems, and the number of inputs and outputs for the boxes of each system. One way to find the set of all extremal effects in a given scenario is using vertex enumeration (all the extremal no-signaling distributions act as facets since the inner product of each such state with any valid effect must be greater than 0 and less than 1). However, directly performing vertex enumeration is slow, except in the smallest scenario [27]. To circumvent this we exploit an alternative method for finding extremal effects that starts with different ways to represent the identity effect. By breaking down these identity effects we can find the extremal effects in various scenarios. We separately consider deterministic and nondeterministic effects and classify the effects into wirings and nonwirings before investigating the significance of the latter for state discrimination and nonlocality distillation, showing the advantages of nonwiring measurements. We also discuss how our findings relate to the phenomenon of quantum nonlocality without entanglement [28], which is the existence of measurements comprising product effects that cannot be performed by LOCC. The nonwiring-type measurements of this paper serve as a box-world analog of these, and we find a set of product states in box world that can be perfectly distinguished with a nonwiring measurement, but not with any wiring.

Understanding measurements in GPTs is also useful to gain insight into which features of quantum theory make it special with respect to other theories, which in turn may help find underlying principles on which quantum theory can be based. Box world is known to have only separable measurements [26] (i.e., those for which all effects can be expressed as a sum of product effects), which in a sense makes it inferior to other GPTs. However, this property also makes box world a

suitable example for studying the difference between separable measurements that can be done using LOCC and those that cannot, and of the information processing capabilities enabled by the latter, an understanding that will likely carry over to arbitrary GPTs. Further insights into this difference are also desirable in the quantum case [29–32].

The rest of the paper proceeds as follows. We first give a short introduction to GPTs, before introducing the notation and technical background in Sec. III. In Sec. IV, we present technical results underlying the algorithms for generating all extremal effects in box world; these algorithms are presented in Sec. V. In Sec. VI we apply our algorithm to various scenarios of up to four inputs and up to four outputs and demonstrate the application of these results in information processing tasks in Sec. VII. Finally, we connect our work to logically consistent classical processes and to composition of contextuality scenarios in Sec. VIII before making a few concluding remarks in Sec. IX.

II. BACKGROUND

We briefly outline the framework for GPTs that will be used throughout this paper. In a GPT, states are represented as vectors in a vector space V . We use $\mathcal{S} \subseteq V$ to denote the set of all possible states (note that the state space depends on the size of the system), and we typically assume \mathcal{S} is convex and compact. An *effect* is a linear map from a state to a probability. Effects can be taken as vectors in the vector space dual to V , which we call \mathcal{E} , and the map from states to probabilities is then formed by taking the inner product between the state and the effect. Given a state space, a valid effect e must satisfy $0 \leq \langle e, s \rangle \leq 1$ for all $s \in \mathcal{S}$. Under the *no restriction hypothesis* [33], which we will assume here, this necessary condition is taken to be necessary and sufficient for an effect to be valid.

The state space of a composite system is formed by taking some kind of tensor product between the individual state spaces. A minimal requirement is that if $s_A \in \mathcal{S}_A$ is a state of system A and $s_B \in \mathcal{S}_B$ is a state of system B , then $s_A \otimes s_B$ is a state of AB . If all joint states have this form, or are convex combinations of states of this form, then the joint state space is said to be the minimal tensor product of the individual state spaces. Taking the composite system effect space to obey the no restriction hypothesis, the min tensor product state space corresponds to forming the effect space by taking the maximal tensor product of the individual effect spaces.

One way to specify a state is via the probabilities of the outcomes of a set of *fiducial measurements* [2], i.e., a set that is sufficient to completely characterize the state. We will work with systems that have a finite set of fiducial measurements, each of which has a finite number of outcomes. This means that a state can be specified using a vector whose entries contain the probabilities of outcomes for the fiducial measurements. For simplicity, we consider cases in which the number of outcomes of each fiducial measurement is the same. For instance, if for a single system there are two two-outcome fiducial measurements, we write the state as

$$(P(0|0), P(1|0) \mid P(0|1), P(1|1)), \quad (1)$$

where $P(i|j)$ is the probability of output i given input j .

¹More precisely, wirings only require one-way LOCC.

Although this is a four-dimensional vector, due to normalization there are only two independent parameters. For later convenience we stick with the larger representation rather than suppressing the redundant parameters.

In this work we assume *local tomography* [34], i.e., that the state of a joint system can be determined from the statistics of local fiducial measurements (called the global state assumption in [8]). This means that a state of N parties (where for each single system there are two two-outcome fiducial measurements) has a similar representation:

$$(P(0\dots 0|0\dots 0), \dots, P(1\dots 1|0\dots 0) | \dots | P(0\dots 0|1\dots 1), \dots, P(1\dots 1|1\dots 1)) \quad (2)$$

(the ordering is such that first for input $00\dots 0$ the outcomes increase counting in binary, and then the inputs increase counting in binary).

For every system there is an identity effect, i.e., an effect u such that $\langle u, s \rangle = 1$ for all $s \in \mathcal{S}$. In our notation there are several ways to write this effect. For a single system we can write $u^R = (1, 1|0, 0)$, $u^R = (0, 0|1, 1)$, $u^R = (1, 1|1, 1)/2$, etc. Although written differently, these all represent the same effect, the notation u^R meaning a representation of the effect u . In general, we identify two vectors as representing the same effect if they have the same inner product with all elements of the state space.

The set of vectors that can be added to any effect vector without changing the effect it represents we term *no-signaling moves* because they each arise as a result of the state space \mathcal{S} only containing no-signaling distributions (or as a result of the normalization). For instance, for all valid states, $P(00|00) + P(10|00) - P(00|10) - P(10|10) = 0$, which represents the impossibility of Alice's choice of measurement affecting Bob's outcome when Bob makes input 0. This no-signaling condition can be encoded using a vector r such that $\langle r, s \rangle = 0$ for all $s \in \mathcal{S}$. Thus, if e^R is a vector representing a valid effect, then $e^R + r$ represents the same effect. We use $\{r_i\}$ to denote a complete set of generators of all such vectors, so that any no-signaling move can be written as a linear combination of vectors from $\{r_i\}$. A *measurement* is a collection of effects that sum to the identity effect, i.e., given state space \mathcal{S} , the set of effects $\{e_1, \dots, e_m\}$ form a measurement on \mathcal{S} if for each $i = 1, \dots, m$ we have $0 \leq \langle e_i, s \rangle \leq 1$ for all $s \in \mathcal{S}$ and $\sum_{i=1}^m e_i = u$.

In this work we will be interested in the effect space when the state space comprises all no-signaling distributions (also known as *box world*). Because box world has large multipartite state spaces, the set of possible effects is comparably restricted. Box-world effects are therefore valid in a wide range of GPTs. We consider the effect spaces for various small numbers of inputs, outputs, and parties (larger cases become too computationally intensive). Due to linearity, the effect space is convex, and can hence be characterized in terms of a set of extremal effects. There are a finite number of these, generating a convex polytope; our aim is to have a procedure that can generate the vertices of this polytope.

In the two-party case, the complete set of extremal effects was computed in [27]. It was found that there are 82 such effects. These were computed by vertex enumeration starting from the facet description of the effect space. This facet

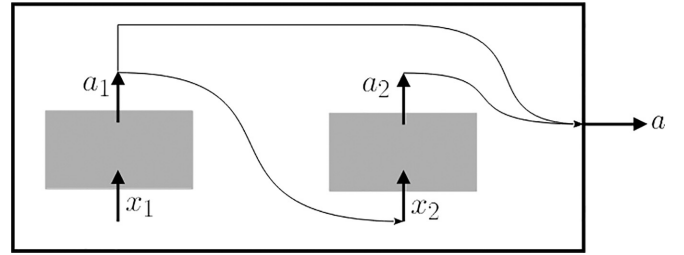


FIG. 1. Illustration of a wiring effect in the bipartite case. An input x_1 is made generating outcome a_1 , then x_2 is taken as a function of a_1 giving outcome a_2 . The final output a is then a function of a_1 and a_2 . For instance, if $x_1 = 0$, $x_2 = a_1$, and $a = a_1 \oplus a_2$ then the effect corresponding to $a = 0$ is $e_{a=0} = (1, 0, 0, 0|0, 0, 0, 1|0, 0, 0, 0|0, 0, 0, 0)$ and the effect corresponding to $a = 1$ is $e_{a=1} = (0, 1, 0, 0|0, 0, 0, 1, 0|0, 0, 0, 0|0, 0, 0, 0)$.

description says that to be a valid effect $0 \leq \langle e, s_i \rangle \leq 1$ for all extremal states s_i (in the present case there are 24 extremal states: 16 local deterministic states and 8 PR-box-type states). In this work we use a different method that allows us to treat cases for which our computational tools for vertex enumeration are prohibitively slow.

We will also be interested in a special type of effect called a *wiring*. Wirings are effects that can be implemented by convex combinations of procedures of the following form: choose one of the states and choose a measurement to make on that state, take its output, and apply a function to it to choose the next state and the measurement on the next state, and so on, where at each step all the previous outputs are used as arguments of the function that selects the next state and the measurement performed on it. The final output is then formed by taking a function of all the individual measurement outcomes (see Fig. 1 for an illustration). The extremal wirings are effects of this type that cannot be represented as convex combinations of others. These box-world measurements are of particular interest since they correspond to classical processing of inputs and outcomes and are thus applicable in any operational theory allowing classical processing. In particular, they are applicable irrespective of the type of system under consideration and can be used, for instance, for nonlocality distillation, which in turn may have applications for device-independent information processing. This is discussed further in Sec. VII C.

III. PRELIMINARIES

Let n_I be the number of inputs and n_O be the number of outputs per party with N parties. A state can then be represented as a $(n_I n_O)^N$ -dimensional vector, whose entries are the conditional probabilities of every set of outputs given every set of inputs. Given a subset V of the set $[N] := \{0, 1, 2, \dots, N-1\}$ of parties, we use X_V as the random variable representing their inputs (elements of $[n_I]^{|V|}$) and A_V as that for their outputs (elements of $[n_O]^{|V|}$).

We will now define our state space \mathcal{S}_{NS} . To be elements of this we require three conditions to hold: no signaling, positivity, and normalization, which we detail below. The

no-signaling condition is that no subset of parties should be able to signal to any other subset of parties. In other words, for

any disjoint subsets V and W of $[N]$, the parties in V cannot signal to those in W . Mathematically, this means

$$\sum_{\mathbf{a}_V} P_{A_W A_V | X_W = \mathbf{x}_W, X_V = \mathbf{0}}(\mathbf{a}_W, \mathbf{a}_V) = \sum_{\mathbf{a}_V} P_{A_W A_V | X_W = \mathbf{x}_W, X_V = \mathbf{x}'_V}(\mathbf{a}_W, \mathbf{a}_V)$$

for all $\mathbf{a}_W \in \{0, 1, 2, \dots, n_O - 1\}^{|W|}$, $\mathbf{x}_W \in \{0, 1, 2, \dots, n_I - 1\}^{|W|}$, and $\mathbf{x}'_V \in \{0, 1, 2, \dots, n_I - 1\}^{|V|}$. (3)

A smaller subset of this set of conditions is sufficient to generate the whole set, namely, it suffices that no party can signal to the collection of all the others. This is stated formally in the following lemma.

Lemma 1. The set of conditions implied by (3) for all disjoint subsets V and W of $[N]$ is implied by demanding that (3) holds for the subset of conditions where V is a singleton and $W = [N] \setminus V$.

The conditions where V is a singleton and $W = [N] \setminus V$ can be represented by a set of vectors $\{r_i\}$, such that $\langle r_i, s^R \rangle = 0$ for all $s \in \mathcal{S}_{NS}$, and where each r_i contains n_O elements with value 1, n_O elements with value -1 , and the remaining entries are 0. The number of such conditions is $N(n_O n_I)^{N-1} (n_I - 1)$. These are not all linearly independent: the dimension of the space of un-normalized no-signaling distributions is $[n_I(n_O - 1) + 1]^N$ [35,36].

A valid state must also have positive entries since each entry represents a probability, i.e.,

$$P_{A_S | X_S}(\mathbf{a} | \mathbf{x}) \geq 0 \text{ for all } \mathbf{a}, \mathbf{x}, \quad (4)$$

where $S = [N]$.

The remaining condition for a vector to be an element of \mathcal{S}_{NS} is that

$$\sum_{\mathbf{a}_S} P_{A_S | X_S = \mathbf{0}}(\mathbf{a}_S) = 1, \quad (5)$$

where $S = [N]$. Note that, given the no-signaling conditions, if (5) holds for $X_S = \mathbf{0} := 0 \dots 0$ then it also holds for $X_S = \mathbf{x}_S$ for any other $\mathbf{x}_S \in [n_O]^{|S|}$. Furthermore, together with (4) this implies that no entry can exceed 1 (so conditions that each probability is at most 1 do not need to be added).

An important set of states are the *local deterministic states*. These correspond to deterministically assigning one of the outcomes for every possible input of every party. Thus, they are states of the form

$$P_{A | X = \mathbf{x}}(\mathbf{a}) = P_{A_1 | X_1 = x_1}(a_1) P_{A_2 | X_2 = x_2}(a_2) \dots P_{A_N | X_N = x_N}(a_N),$$

and where $P_{A_i | X_i} \in \{0, 1\}$ in all cases. We denote these $s^{L,i}$, where i runs from 1 to $(n_I n_O)^N$. The linear span of the set of local deterministic states is the same as that of the no-signaling distributions (being local does not confer any additional *equality* constraints over being no signaling). Thus, given two vectors representing effects, we can check whether they represent the same effect by checking whether both vectors have the same inner product with all local deterministic distributions.²

An important symmetry class for a given state or effect is that of the relabelings. We can relabel the parties, the inputs for each party, or the outputs for each input of each party. Thus, the total size of the symmetry group of relabelings is $[(n_O!)^{n_I} n_I!]^N N!$.

IV. EXTREMAL EFFECTS

We are interested in finding the extremal effects for a given scenario, i.e., for a given (N, n_I, n_O) . As mentioned before we use a representation that has some redundancy in that the constraints arising from the no-signaling moves and normalization are not used to reduce the parameters. (Such a reduction would typically be done when trying to solve this problem directly by means of vertex enumeration.) Instead, the technical results of this section show that the redundant representation employed has a convenient structure that allows us to find extremal effects in a different way, which will be the basis of the algorithms we present in Sec. V. We start with the following observation, expressed in terms of a *standard basis vector* (SBV), by which we mean a vector with all components 0 except for one 1.

Lemma 2. Every $(n_I n_O)^N$ -dimensional SBV is a representation of an extremal effect.

Proof. To see that SBVs all represent effects, note that the inner product of such a vector e^R with a state $s \in \mathcal{S}$ gives a single probability, hence, $0 \leq \langle e^R, s^R \rangle \leq 1$ for all $s \in \mathcal{S}$, making e^R a representation of a valid effect. By considering a state satisfying $\langle e, s \rangle = 1$, it is clear that αe is not an effect for any $\alpha > 1$, hence, e is on the boundary of the effect space. To see that it is extremal, suppose $e = \alpha e_1 + (1 - \alpha) e_2$ for two effects e_1 and e_2 and $0 < \alpha < 1$. Then

$$\langle e, s \rangle = \alpha \langle e_1, s \rangle + (1 - \alpha) \langle e_2, s \rangle.$$

Now suppose s_1 is a state with $\langle e, s_1 \rangle = 1$. Since e_1 and e_2 are effects, it follows that $\langle e_1, s_1 \rangle = \langle e_2, s_1 \rangle = 1$. Similarly, if s_0 is a state with $\langle e, s_0 \rangle = 0$, it follows that $\langle e_1, s_0 \rangle = \langle e_2, s_0 \rangle = 0$. Because the local deterministic distributions are $\{0, 1\}$ valued, if s is local deterministic it satisfies $\langle e, s \rangle \in \{0, 1\}$. It follows that both e_1 and e_2 have the same action on any local deterministic distribution as e does. Since the set of local deterministic distributions span the space of no-signaling distributions, e_1 and e_2 must have the same action as e for any $s \in \mathcal{S}_{NS}$. Thus, e_1, e_2 , and e must be the same effect and e is extremal. ■

The $\{0, 1\}$ -valued representations of the identity effect will be important because these are the class that can be used to generate all wirings (as well as some nonwirings).

Lemma 3. Let u^R be a $\{0, 1\}$ -valued vector representing the identity effect. Any vector e^R formed from u^R by replacing

²In general, a subset of these would also suffice.

any number of the 1 entries with 0s is a representation of an extremal effect.

Proof. First, since u^R represents the identity effect we have $\langle u^R, s^R \rangle = 1$ for all $s \in \mathcal{S}$. Since u^R is a $\{0, 1\}$ -valued vector we can write it in terms of the SBV effects $\{e_i\}$ as $u = \sum_i \lambda_i e_i$, where $\lambda_i \in \{0, 1\}$. We can also write $e^R = \sum_i \lambda'_i e_i^R$, where $\lambda'_i \in \{0, 1\}$ and $\{i : \lambda'_i = 1\} \subset \{i : \lambda_i = 1\}$. It follows that $0 \leq \langle e, s \rangle \leq \langle u, s \rangle = 1$ for all $s \in \mathcal{S}$, so e is a valid effect.

To see that it is extremal, consider writing $e = \alpha e_1 + (1 - \alpha) e_2$ for two effects e_1 and e_2 and $0 < \alpha < 1$. Since e^R is a $\{0, 1\}$ -valued representation of e , and all local deterministic states have a $\{0, 1\}$ -valued representation, for a local deterministic state s we have $\langle e, s \rangle \in \{0, 1\}$. Hence, by the same argument as in Lemma 2, e must be extremal. ■

Lemma 4. Any extremal wiring has a representation that can be formed by taking the $\{0, 1\}$ -valued representation of the identity effect corresponding to all parties making the input 0, applying no-signaling moves in such a way that it remains $\{0, 1\}$ valued, and replacing some of the entries that are 1 with 0.

Proof. Consider first the case $N = 1$. In this case the extremal wirings are the effects formed by choosing one of the n_I measurements and then applying a deterministic function from $\{0, 1, \dots, n_O - 1\}$ to $\{0, 1, \dots, n_O - 1\}$ to the outcome. In this case the no-signaling moves³ take us from the identity for input 0 to that for all other choices of measurement so the statement holds.

Now assume by induction that the statement holds for $N - 1$ parties and consider N parties. The first step in an extremal wiring is to choose one of the N boxes and make a fixed input to that box. Up to symmetry, we can assume the first box is chosen and input 0 is made (relabeling parties or inputs does not change whether an effect is a wiring or not). In this case, the effect can only have nonzero entries where these correspond to elements of the state of the form $P(a_1 a_2 \dots a_{n_I} | 0 x_2 \dots x_{n_I})$, i.e., where $x_1 = 0$. For each outcome we can then consider the $N - 1$ party effect that is performed conditioned on $x_1 = 0$ and the value of a_1 . By assumption, each of these subeffects can be formed by taking the $\{0, 1\}$ -valued representation of the identity effect corresponding to $N - 1$ parties making the input 0, applying no-signaling moves in such a way that it remains $\{0, 1\}$ valued, and replacing some of the entries that are 1 with 0. Let us ignore the replacement of 1 entries with 0s for the moment and consider only the representation of identity. Up to no-signaling moves each of the subeffects corresponds to all parties measuring 0, and, up to symmetry the measurement on the first box corresponds to $x_1 = 0$. Thus, up to no-signaling moves, the identity is that corresponding to all parties measuring 0, and the effect is then formed by zeroing entries of this. ■

Lemma 5. Every effect can be represented by a $(n_I n_O)^N$ -dimensional vector in which every entry is non-negative.

Proof. This is proven as part of Theorem 7 in [8]. We give the argument for completeness. Consider the cone of non-normalized states in our representation, i.e., the set of

$(n_I n_O)^N$ -dimensional vectors

$$\mathcal{V} = \{v : \langle e_i, v \rangle \geq 0 \forall i, \langle r_j, v \rangle = 0 \forall j\},$$

where $\{e_i\}$ are the SBV effects and r_j are vectors representing the no-signaling moves. We can rewrite $\langle r_j, v \rangle = 0$ as $\langle r_j, v \rangle \geq 0$ and $\langle -r_j, v \rangle \geq 0$. The dual cone is then that formed by the conic hull of $\{e_i\}_i \cup \{r_j\}_j \cup \{-r_j\}_j$. Thus, any effect can be written as $\sum_i t_i e_i + \sum_j w_j r_j$, where $t_i \geq 0$, but w_j can be negative. Since $\langle r_j, v \rangle = 0$ for all $v \in \mathcal{V}$, one representation of the effect is when the values of $\{w_j\}$ are set to zero. Thus, any effect can be written in the form $\sum_i t_i e_i$ where $t_i \geq 0$. ■

It is helpful to consider the set of $(n_I n_O)^N$ -dimensional identity effects that have positive entries. These form a convex polytope since they are defined by the vectors u^R for which every element is positive and such that $\langle s^{L,i}, u^R \rangle = 1$ where i runs over all local deterministic distributions. There are hence $(n_I n_O)^N$ equality constraints and $(n_I n_O)^N$ inequality constraints. We call the extreme points of this polytope the *extremal representations of the identity effect*, and these can be computed using vertex enumeration.

Lemma 6. Every extremal effect has a representation as a vector that can be formed by taking an extremal representation of the identity effect and replacing some of the nonzero entries with 0s.

Proof. Let e be an effect and $f = u - e$. By Lemma 5, we can represent e and f using $(n_I n_O)^N$ -dimensional vectors e^R and f^R whose entries are non-negative. Write $e^R = \sum_i \lambda_i e_i$ and $f^R = \sum_i \mu_i e_i$, where $\{e_i\}$ are the SBV effects and $\{\lambda_i\}$ and $\{\mu_i\}$ are non-negative. Thus, $u^R = \sum_i (\lambda_i + \mu_i) e_i$ is a representation of the identity effect.

We claim that if e is extremal, then for each i either $\lambda_i = 0$ or $\mu_i = 0$. Suppose by contradiction that there is some j for which $0 < \lambda_j < \lambda_j + \mu_j$. Then,

$$\begin{aligned} e^R &= \lambda_j e_j + \sum_{i \neq j} \lambda_i e_i = \frac{\lambda_j}{\lambda_j + \mu_j} \left((\lambda_j + \mu_j) e_j + \sum_{i \neq j} \lambda_i e_i \right) \\ &\quad + \left(1 - \frac{\lambda_j}{\lambda_j + \mu_j} \right) \sum_{i \neq j} \lambda_i e_i \\ &= \frac{\lambda_j}{\lambda_j + \mu_j} (e^R + \mu_j e_j) + \left(1 - \frac{\lambda_j}{\lambda_j + \mu_j} \right) (e^R - \lambda_j e_j). \end{aligned}$$

Both $e^R + \mu_j e_j$ and $e^R - \lambda_j e_j$ represent effects [for the former, note that $e^R + \mu_j e_j = u^R - (f^R - \mu_j e_j)$]. Hence, we have decomposed e as a convex combination of other effects, contradicting the assumption that e is extremal. Note that this implies that for any extremal effect, there exists a representation of it and its complement that are orthogonal.⁴ Hence, we have shown that if e is extremal, it can be formed by zeroing entries from a representation of an identity effect. It remains to show that nonextremal representations of the identity effect need not be used.

³The term no signaling does not make sense for $N = 1$, but the mathematical conditions are well defined and take the form $\sum_a P_{A|X=x}(a) = \sum_a P_{A|X=x'}(a)$ for all x, x' .

⁴In fact, what we have shown is even stronger: there is a representation in which the elementwise product of e^R and f^R is zero.

Suppose u^R is a nonextremal representation of the identity effect, so $u^R = \sum_i v_i u_i^R$ with $\{u_i^R\}$ being extremal representations of the identity effect, $v_i \geq 0$, $\sum_i v_i = 1$ and at least two $v_i > 0$. Let Z be a map that zeroes some of the entries and suppose $e^R = Z(u^R) = \sum_i v_i Z(u_i^R)$. If there are two or more values of i for which both $v_i > 0$ and $Z(u_i^R) \neq 0$, then e^R is not extremal. If there is only one i such that both $v_i > 0$ and $Z(u_i^R) \neq 0$, then e^R is also not extremal (but is proportional to the zeroing of an extremal representation of the identity effect). ■

V. COMPUTING EXTREMAL EFFECTS

Our method for computing extremal effects is suggested by Lemma 6. The first step is to identify all extremal representations of the identity effect by means of a vertex enumeration. Specifically, the condition $\langle u, s \rangle = 1$ for all elements of the state space can be imposed by requiring $\langle u, s \rangle = 1$ for all local deterministic s since the linear span of the local deterministic states covers the state space. This means that the set of identity effects for N parties can be expressed using 4^N equality constraints and 4^N inequality constraints (positivity of the individual entries). This problem is more tractable than performing the full vertex enumeration to compute all extremal effects directly. Instead, the extremal effects are obtained from the identity effects by considering subeffects (see Lemma 6).

Computing the set of all extremal effects scales badly with the parameters of the scenario and hence we do not compute all in most cases (although our algorithm would in principle allow this). In several scenarios we instead compute the deterministic extremal effects, which include all extremal wirings.

A. Computing deterministic extremal effects

Our method for computing all deterministic extremal effects is suggested by Lemma 3. This algorithm recovers all deterministic extremal effects, which includes all wirings (cf. Lemma 4).

We first find all the $\{0, 1\}$ -valued representations of the identity effect. This can be done by the following algorithm:

Algorithm 1: generate $\{0, 1\}$ -valued identity effects

(1) Let $S = \{u_1^R\}$ where u_1^R is any $\{0, 1\}$ -valued representation of the identity effect. [The algorithm can also be started with any initial set of $\{0, 1\}$ -valued representations of the identity effect.]

(2) Generate $S' = \{s_j \pm r_i\}_{i,j}$, where s_j are elements of S and r_i are a complete set of generators of the no-signaling moves taking values $\{-1, 0, 1\}$ (cf. Lemma 1).

(3) Remove elements of S' with negative entries and set $S = S'$.

(4) Repeat Steps (2) and (3) until S stops increasing.

(5) Output S .

Algorithm 1 is a subalgorithm of our main routine:

Algorithm 2: generate all $\{0, 1\}$ -valued effects

(1) Use Algorithm 1 to generate all representations of the identity effect.

(2) For each representation of the identity, form a new set of effects containing all effects that are obtained by deleting any number of 1s from each of the representations.

(3) Take the union of all the sets generated.

(4) For each element e_i in this union, compute $(e_i, L(e_i))$, where $L(e_i) = (\langle e_i, s^{L,1} \rangle, \langle e_i, s^{L,2} \rangle, \dots)$ and generate a list $S = ((e_1, L(e_1)), (e_2, L(e_2)), \dots)$ of all these pairs.

(5) Go through the list checking whether $L(e_i) = L(e_j)$ where $i \neq j$. If so, remove one of the two elements from S .

(6) Output S .

Note that each identity effect has n_O^N elements with value 1, so there are $2^{n_O^N}$ ways of deleting 1s for each representation of the identity effect in Step (2) (hence this algorithm does not scale well as the number of parties increases). Taking the union of the sets involves removing any duplicate representations. However, at this point there remain different representations of the same effect in our set. The use of $L(e_i)$ is a convenient way to remove such different representations.

We are also interested in classifying the extremal effects as either wirings or nonwirings. We first classify a set of *wiring representations* using the following algorithm (cf. Lemma 4):

Algorithm 3: classify a representation of an effect as wiring representation

This algorithm takes as input a representation e^R of an effect.

(1) If the number of parties is 1, return 1.⁵

(2) Run over all exchanges of party 1 with each other party, and exchanges of input labels for the chosen party until the resulting effect (after the exchanges) has zero elements wherever $x_1 \neq 0$. If no such case is found, return 0.

(3) For each of the possible outcomes a_1 compute the $N - 1$ party effect conditioned on $x_1 = 0$ and the value of a_1 (there are n_O instances to compute).

(4) Recursively run the same algorithm on each of these $N - 1$ party effects. If all cases return 1 then return 1, otherwise return 0.

Algorithm 3 outputs 0 if its input is not a wiring representation and outputs 1 if it is. An effect is a wiring if and only if it can be expressed as a convex combination of effects that have wiring representations. An extremal effect is hence a wiring if and only if it has a wiring representation.

To connect Algorithm 3 to the previously mentioned notion of a wiring, consider the first step in an extremal wiring. This involves choosing one of the boxes to make an input to as well as the value of the input. Consider the case where this is the first box and the input made is $x_1 = 0$. In this case, e^R will have zeros for the elements corresponding to any probabilities conditioned on other values of x_1 . Hence, if e^R is a wiring, there must exist a permutation Π of parties and of input labels such that Πe^R has zero elements wherever $x_1 \neq 0$. For each of the possible outcomes a_1 we can consider the $N - 1$ party effect conditioned on $x_1 = 0$ and the value of a_1 (that each of these are an effect follows by considering the set of states that are a tensor product of the deterministic state that always outputs a_1 for $x_1 = 0$ with any $N - 1$ party state on the remaining systems). We can check that all these smaller effects have an analogous property in the same way and recurse. If the required label permutation exists at all levels, we can conclude that the effect corresponds to a wiring.

⁵See later for an explanation as to why we can increase the number of parties to 2 in this step.

Suppose we run Algorithm 3 on a particular $\{0, 1\}$ -valued representation e^R of an effect e . If we get output 1 then we know that e can be implemented as a wiring. However, if we get 0, it could be that there is an alternative representation of e that is a wiring representation. To understand which effects are wirings or not we can modify Steps (4) and (5) of Algorithm 2 to the following:

(4') For each element e_i in this union, compute $(e_i, L(e_i), W(e_i))$, where $L(e_i)$ is as before and $W(e_i)$ is whether the representation is a wiring representation or not and generate a list $S = (e_1, L(e_1), W(e_1)), (e_2, L(e_2), W(e_2)), \dots)$ of all these pairs.

(5') Go through the list checking whether $L(e_i) = L(e_j)$ where $i \neq j$. If so, check whether $W(e_i) = 1$. If so, remove element j from the list, otherwise remove element i .

With this modified algorithm, we can identify effects as either wirings or nonwirings as they are found.

VI. RESULTS

A. Two inputs and two outputs per party

In this section we discuss cases with $n_I = 2$ and $n_O = 2$ for $N = 2, 3, 4$.

1. $N = 2$

This case was already computed in [27] using a different procedure. The output of our algorithm in this case agrees with that of [27]. In particular, there are 82 extremal effects. We break these down into 7 classes (two effects are in the same class if they are equivalent up to relabelings, where we can relabel the parties, the inputs for each party, and the outputs for each input of each party).⁶ All of the effects are wirings, which was already known from [26]. That all of the effects are wirings means we can alter Step (1) of Algorithm 3 to “if the number of parties is 2, return 1.”

In this case the deterministic effects are sufficient for describing the full effect polytope, so the more general case is omitted here.

2. $N = 3$

A complete description of all deterministic effects was not known for this case, although in [26] it was noted that for three parties not all effects are wirings. We have computed all the $\{0, 1\}$ -valued extremal effects for this case, finding 28 886 effects of which 2048 are nonwirings. These extremal effects break down into 66 classes of which 3 correspond to nonwirings. They are collected in the online Supplemental Material [37]. Representatives of each of the classes of nonwirings are as follows:

$$\langle e_1, s \rangle = P(000|100) + P(001|010) + P(110|001),$$

$$\langle e_2, s \rangle = P(000|100) + P(001|010) + P(110|001) \\ + P(010|000),$$

$$\langle e_3, s \rangle = P(000|100) + P(001|010) + P(110|001) \\ + P(010|000) + P(101|000).$$

Performing a vertex enumeration using the software PORTA [38], we are also able to find all the extremal representations of the identity effect. It turns out that there are 710 760 of these, of which 744 are $\{0, 1\}$ valued (and 680 of the latter are wiring representations). These extremal representations of the identity effect break down into 307 classes, of which 9 are $\{0, 1\}$ valued (and 8 of the latter are wiring representations). Representatives of each of the 307 classes can be found in the Supplemental Material [37].

In principle, these representations can be used to generate all extremal effects as suggested by Lemma 6. However, this computation is prohibitively time consuming because many of the classes of extremal representations of the identity effect have 27 nonzero entries, and for these there are 2^{27} candidate effects that can be formed by zeroing various entries (by contrast, the $\{0, 1\}$ -valued extremal representations of the identity effect have only 8 nonzero entries, which is why we could solve this case). Nevertheless, to illustrate that this can in principle be done we have taken two of the identity effects with a smaller number of nonzero entries, namely, u_1^R and u_2^R satisfying

$$\langle u_1^R, s \rangle = \frac{1}{3}(P(000|101) + 2P(000|110) + P(001|010) + P(001|011) + P(001|100) + P(010|010) + P(010|011) \\ + P(010|101) + P(011|100) + 2P(011|111) + P(100|100) + 2P(100|111) + P(101|010) + P(101|011) \\ + P(101|101) + P(110|010) + P(110|011) + P(110|100) + P(111|101) + 2P(111|110)),$$

$$\langle u_2^R, s \rangle = \frac{1}{2}[P(000|011) + P(000|101) + P(001|011) + P(001|101) + P(010|101) + P(010|110) + P(011|010) \\ + P(011|101) + P(100|011) + P(100|100) + P(101|011) + P(101|110) + P(110|100) + P(110|111) \\ + P(111|010) + P(111|111)],$$

and computed all the subeffects from these (up to symmetry), using a linear program to remove any that are convex combinations of the extremal deterministic effects we already found. The output of this computation is given in the Supplemental Material

⁶Note that in [27] they classify differently into 5 classes.

[37]. The following are two examples:

$$\begin{aligned}\langle e_4, s \rangle &= \frac{1}{3}[P(001|010) + P(001|100) + P(010|011) + P(101|010) + P(111|101) + 2P(111|110)] \\ \langle e_5, s \rangle &= \frac{1}{2}[P(000|011) + P(001|011) + P(010|110) + P(011|010) + P(100|100) + P(111|010)].\end{aligned}$$

3. $N = 4$

In this case computing all the extremal representations of the identity effect is not feasible in reasonable time using PORTA. Furthermore, the number of extremal effects is too large to directly use our previous technique for the $\{0, 1\}$ -valued extremal effects. Instead, we can compute all the $\{0, 1\}$ -valued extremal effects by computing one representative of each symmetry class. The computation works in the same way as before, but we remove symmetries at every step.

In particular, in Algorithm 1 we add a step between Steps (3) and (4) that removes elements of S that are equal to others under relabeling symmetries. In Algorithm 2, rather than using the list of local values $L(e_i)$ we generate a canonical form of these by generating $L(e_i)$ for every symmetry of e_i and then storing the first of all of these according to some ordering function [e.g., since each list $L(\Pi e_i)$ is $\{0, 1\}$ valued, they can be ordered as a binary number]. We run Algorithm 2 with the modification to classify into wiring or nonwiring representations (i.e., using Steps (4') and (5')).

Overall we find 168 301 classes of extremal $\{0, 1\}$ -valued effect, of which 124 698 are wiring representations. By generating all the symmetries of each, we can then compute the total number of $\{0, 1\}$ -valued effects to be 7 940 781 474, of which 4 729 832 866 are wiring representations. Because of the size, we only supply supplemental files with an element of each class in this case [37].

B. Generalizations: More inputs and outputs

In the cases $n_I = 2$ and $n_O = 2$ we were only able to partially solve the cases with $N = 3$ and 4. Increasing the number of inputs and outputs further increases the complexity, but we can make a few remarks. First, consider the case $N = 2$. It was proven in [26] that the bipartite effect spaces also only contain wirings. Using our code we enumerate the number of classes for the first few cases, as well as the total number of effects (see Table I). Data with the full set of extremal effects for these cases can be found in the Supplemental Material [37].

TABLE I. Numbers of classes and effects for two parties with various numbers of inputs and outputs. (We use a dash to indicate that we did not compute this case). The number of classes in the case $n_O = 2$ will remain at 7 for any n_I .

n_O	2	3	4
Classes $n_I = 2$	7	44	523
Classes $n_I = 3$	7	48	—
Classes $n_I = 4$	7	—	—
Total $n_I = 2$	82	8930	2977858
Total $n_I = 3$	248	43400	—
Total $n_I = 4$	562	—	—

In the case $N = 3$, the only additional case we attempt is $n_I = 3, n_O = 2$. Here we use the previous method to compute the $\{0, 1\}$ -valued effects, obtaining 79 classes of such effect of which 76 are wirings and 3 are nonwirings. In total the number of $\{0, 1\}$ -valued effects is 505 136 which breaks down as 449 840 wirings and 55 296 nonwirings. Again, these cases can be found in the Supplemental Material [37].

Our codes can also be used for the enumeration of all extremal effects in these scenarios (within the computational limitations). We omit explicit characterizations here.

VII. APPLICATIONS OF NONWIRINGS

In this section we discuss some applications of nonwiring measurements, focusing on those that outperform wirings.

A. State discrimination

Given a black box that outputs one of two possible (known) states s_1 and s_2 , with probability $1/2$ each, the task is to choose a measurement that gives the highest probability of correctly guessing which state was produced given just one copy. When trying to discriminate between two probability distributions P_X and Q_X , the guessing probability is

$$\frac{1}{2}[1 + D_C(P_X, Q_X)],$$

where D_C is the total variation distance (the subscript C indicating classical), i.e., $D_C(P, Q) = \frac{1}{2} \sum_x |P_X(x) - Q_X(x)|$. In the case of two quantum states ρ_1 and ρ_2 , this optimal guessing probability is

$$\begin{aligned}\frac{1}{2} \max_{E_1, E_2} [1 + D_C(\{\text{tr}(E_1 \rho_1), \text{tr}(E_2 \rho_1)\}, \{\text{tr}(E_1 \rho_2), \text{tr}(E_2 \rho_2)\})] \\ = \frac{1}{2}[1 + D_Q(\rho_1, \rho_2)],\end{aligned}$$

where $\{E_1, E_2\}$ form a positive operator-valued measure (POVM) and D_Q is the trace distance (the subscript Q indicating quantum), i.e., $D_Q(\rho_1, \rho_2) = \frac{1}{2} \text{tr}|\rho_1 - \rho_2|$ (see, e.g., [39,40]).

The analogous formula for box world is that the optimal probability is

$$\begin{aligned}\frac{1}{2} \max_{e_1, e_2} [1 + D_C(\{\langle e_1, s_1 \rangle, \langle e_2, s_1 \rangle\}, \{\langle e_1, s_2 \rangle, \langle e_2, s_2 \rangle\})] \\ = \frac{1}{2} \max_{e_1} (1 + |\langle e_1, s_1 \rangle - \langle e_1, s_2 \rangle|),\end{aligned}$$

where the first maximization is over all measurements $\{e_1, e_2\}$ and the second is over all effects e_1 . The optimum will always be achieved by an extremal effect, hence, in cases where we have computed all extremal effects, we can calculate it by running over all of these. By analogy with the quantum and classical cases, it is natural to define $D_B(s_1, s_2) := \max_{e_1} |\langle e_1, s_1 \rangle - \langle e_1, s_2 \rangle|$ (it is not clear how to remove the maximization from this expression in this case). The quantity D_B also satisfies the requirements of a distance measure [41].

inequality between a vector and a number is interpreted elementwise.) A nonwiring e is advantageous if $v < \mu$. In the case of perfect distinguishability, μ is set to 1. In order to find any separation, μ could be taken as a variable that is optimized and $\mu - v$ maximized instead.

We have implemented this program in MATLAB, relying on YALMIP [42] and MOSEK [43] to solve the linear programs. Using this program, various examples analogous to the ones above can be found. Checking all effects mentioned in Sec. VIA 2 with this program, we find that there are also examples of nonwirings that only allow for distinguishing states perfectly that can also be perfectly distinguished with wirings. In addition, many of the effects outperform wirings with respect to the distinguishing probabilities they achieve for some states, but without reaching perfect distinguishability.

B. Nonlocality without entanglement

The phenomenon of *quantum nonlocality without entanglement* [28] describes the situation in which a set of product effects are combined to make a measurement that cannot be implemented by local operations and classical communication. In a sense these measurements are the quantum analog of nonwirings. Analogously to [28], we construct a set of eight states and an eight-outcome measurement that can perfectly distinguish these, while they are not perfectly distinguishable with wirings. Our states are

$$\begin{aligned} s_1 &= t_3 \otimes t_4 \otimes t_2, & s_2 &= t_1 \otimes t_4 \otimes t_2, & s_3 &= t_4 \otimes t_2 \otimes t_3, \\ s_4 &= t_4 \otimes t_2 \otimes t_1, & s_5 &= t_1 \otimes t_1 \otimes t_3, & s_6 &= t_1 \otimes t_3 \otimes t_1, \\ s_7 &= t_3 \otimes t_1 \otimes t_4, & s_8 &= t_3 \otimes t_1 \otimes t_2, \end{aligned}$$

where $t_1 = (1, 0 \mid 1, 0)$, $t_2 = (1, 0 \mid 0, 1)$, $t_3 = (0, 1 \mid 1, 0)$, $t_4 = (0, 1 \mid 0, 1)$ following the notation from (1). These can be perfectly distinguished by the measurement defined by the effects that for any state s behave as

$$\begin{aligned} \langle e_1, s \rangle &= P(110|000), & \langle e_2, s \rangle &= P(011|001), \\ \langle e_3, s \rangle &= P(111|010), & \langle e_4, s \rangle &= P(100|100), \\ \langle e_5, s \rangle &= P(001|000), & \langle e_6, s \rangle &= P(010|001), \\ \langle e_7, s \rangle &= P(101|010), & \langle e_8, s \rangle &= P(000|100). \end{aligned}$$

To show that no wiring can achieve perfect distinguishability of these eight states, we have enumerated all deterministic eight-outcome wiring measurements. We find that the maximum guessing probability achievable with a wiring when the states are chosen uniformly is $7/8$. This is achieved by the wiring with effects satisfying¹⁰

$$\begin{aligned} \langle e_1, s \rangle &= P(110|000), & \langle e_2, s \rangle &= P(010|000), \\ \langle e_3, s \rangle &= P(111|010), & \langle e_4, s \rangle &= P(100|100), \\ \langle e_5, s \rangle &= P(001|000), & \langle e_6, s \rangle &= P(011|000), \\ \langle e_7, s \rangle &= P(101|010), & \langle e_8, s \rangle &= P(000|100). \end{aligned}$$

¹⁰This corresponds to input $x_3 = 0$, if $a_3 = 0$, input $x_2 = 0$ then $x_1 = a_2 \oplus 1$, while if $a_3 = 1$, input $x_1 = 0$ then $x_2 = a_1$.

Our example disproves Observation 1 from [44]. Indeed we show that these eight states cannot be perfectly distinguished with one-way LOCC in box world, while a global measurement achieves this. We see this as an indication that it is not the local indistinguishability within pairs of local states that causes this phenomenon in this example but rather the existence of separable but global measurements.

Note that, on the other hand, an analog to the bipartite (nine-state) example from [28,45] cannot be constructed since all bipartite measurements in box world (even in higher dimensions) are wirings [26]. In box world all measurements are separable so a distinction between separable and entangled measurements cannot be made. That we can still demonstrate that separable measurements outperform wiring measurements suggests that in some contexts, comparing separable and wiring measurements may be more natural than comparing separable and entangled measurements, although in others, e.g., when considering teleportation, entangled measurements are required.

C. Nonlocality distillation

Consider two parties, Alice and Bob, who hold parts of t bipartite systems, with each subsystem having two inputs and two outputs. For simplicity, take these t systems to be identical (with state \hat{s}). A nonlocality distillation protocol seeks to use these t systems to give a larger violation of a Bell inequality than is possible with only 1. The most general strategy is for each party to associate a t -system two-outcome measurement with each possible input. Such measurements have the form $\{e, u - e\}$ and so can be expressed in terms of one effect. Thus, the overall strategy can be expressed using four effects that act on t systems (one for each of Alice's inputs and one for each of Bob's inputs). Because the individual states are identical, the overall starting state is the t -fold tensor product $s = \hat{s}^{\otimes t}$. If e_x are the effects associated with outcome 0 when Alice's input is x , and f_y are likewise those for Bob, then the outcome probabilities are given by $P'(00|xy) = \langle e_x \otimes f_y, s \rangle$, $P'(01|xy) = \langle e_x \otimes (u - f_y), s \rangle$, etc., where the tensor factors need to be matched appropriately ($s \in \mathcal{S}_{A_1 B_1 A_2 B_2 \dots A_t B_t}$, $e_x \in \mathcal{E}_{A_1 A_2 \dots A_t}$, and $f_y \in \mathcal{E}_{B_1 B_2 \dots B_t}$). That the overall effect is a tensor product reflects the independence of Alice's and Bob's operations. The idea of nonlocality distillation is to choose the four effects e_0, e_1, f_0 , and f_1 so as to maximize the violation of a Bell inequality in the resulting distribution $P'(ab|xy)$. Figure 2 depicts this intuitively for three shared systems.

For systems with two inputs and two outputs, the only extremal Bell inequality (up to symmetry) is the Clauser-Horne-Shimony-Holt (CHSH) inequality, which we can express as $\text{CHSH}[P(ab|xy)] = E_{00} + E_{01} + E_{10} - E_{11}$, with $E_{xy} = P(a = b|xy) - P(a \neq b|xy)$. We hence use this as our measure of nonlocality. Since optimizing over all effects for one party is a linear program, we can run over all extremal effects for Alice and do a linear program for Bob to determine the optimal strategy for CHSH-value distillation given three shared systems [46] (in principle we could also do four, but, given the number of effects, the computation time is prohibitive).

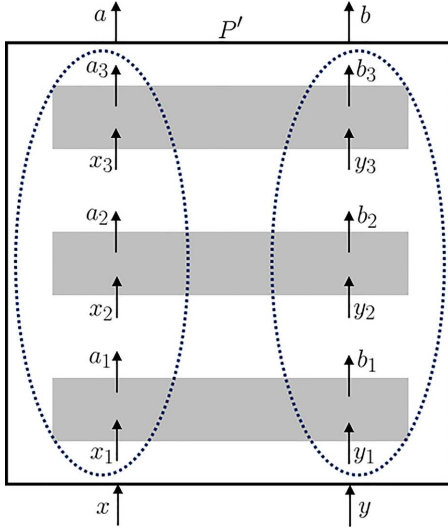


FIG. 2. Illustration of nonlocality distillation in the case $t = 3$. The gray boxes represent the identical initial bipartite states. The outer black frame represents the final correlations P' . The dashed ovals indicate the systems the measurements are performed on (left for Alice; right for Bob).

We perform the optimization for states taken from two two-dimensional cross sections of the set of nonsignaling distributions. To describe these cross sections, we use $P_i^L(ab|xy) = \delta_{a,\mu x \oplus v} \delta_{b,\sigma y \oplus \tau}$ for $\mu, v, \sigma, \tau \in \{0, 1\}$, $i = 1 + \tau + 2\sigma + 4v + 8\mu$, to enumerate the set of 16 deterministic boxes, and $P_i^{\text{NL}}(ab|xy) = \frac{1}{2} \delta_{a \oplus b, xy \oplus \mu x \oplus v y \oplus \sigma}$ for $\mu, v, \sigma \in \{0, 1\}$, $i = 1 + \sigma + 2v + 4\mu$, to denote the set of eight extremal no-signaling boxes [1,35]. In terms of these, the cross sections we have examined are

$$\begin{aligned} \text{CS I} : & \omega P_1^{\text{NL}} + \frac{\eta}{2}(P_1^L + P_6^L) + (1 - \omega - \eta)P^O, \\ \text{CS III} : & \omega P_1^{\text{NL}} + \frac{\eta}{2}(P_1^L + P_9^L) + (1 - \omega - \eta)P^O, \end{aligned} \quad (8)$$

where $P^O = 3/4 P_1^{\text{NL}} + 1/4 P_2^{\text{NL}}$ is local and $\eta, \omega \geq 0$ with $\eta + \omega \leq 1$. The labeling of these cross sections is chosen to follow [46], and the reason we use them is because they have been considered previously.

For the case $t = 3$, the effects found to give rise to CHSH distillation are given below (expressed in terms of their inner product with an arbitrary tripartite state, represented by the ovals in Fig. 2). On Alice's side

$$\begin{aligned} \langle e_0, s \rangle &= P(000|000) + P(011|000) + P(101|000) \\ &\quad + P(110|000), \\ \langle u - e_0, s \rangle &= P(001|000) + P(010|000) + P(100|000) \\ &\quad + P(111|000), \\ \langle e_1, s \rangle &= P(111|011) + P(100|101) + P(001|110) \\ &\quad + P(010|111), \\ \langle u - e_1, s \rangle &= P(011|011) + P(110|101) + P(000|110) \\ &\quad + P(101|111), \end{aligned} \quad (9)$$

while on Bob's

$$\begin{aligned} \langle f_0, s \rangle &= P(011|011) + P(110|101) + P(000|110) \\ &\quad + P(101|111), \\ \langle u - f_0, s \rangle &= P(111|011) + P(100|101) + P(001|110) \\ &\quad + P(010|111), \\ \langle f_1, s \rangle &= P(101|000) + P(000|001) + P(110|010) \\ &\quad + P(011|100), \\ \langle u - f_1, s \rangle &= P(010|000) + P(001|001) + P(100|010) \\ &\quad + P(111|100). \end{aligned} \quad (10)$$

Note that e_1 , f_0 , and f_1 are not wirings. Referring again to Fig. 2, three systems characterized by identical states, each corresponding to coordinates (η, ω) in CS I or CS III and initial nonlocality $\text{CHSH}_i = 2(1 + \omega)$, get distilled through the protocol consisting of the effects (9) and (10), and the final nonlocality (that of the P' correlations) is

$$\text{CHSH}_f^I = \frac{1}{32}[7\omega^3 - 15\eta^3 + 33\omega^2 + 57\omega + 3\eta^2(7 + 11\omega) + 3\eta(9 + 26\omega + 13\omega^2) + 31], \quad (11)$$

$$\text{CHSH}_f^{\text{III}} = \frac{1}{32}[7\omega^3 + 5\eta^3 + 33\omega^2 + 57\omega + \eta^2(13 + 25\omega) + 3\eta(5 + 18\omega + 9\omega^2) + 31]. \quad (12)$$

The distillable region (where $\text{CHSH}_f > \text{CHSH}_i$) for the protocol is shown in Fig. 3 for the two cross sections, CS I and CS III.

D. Limitations for information processing with wirings and box world's nonwiring operations

Despite the advantages we managed to demonstrate above, access to nonwirings in box world does not unlock the same potential as access to measurements that cannot be implemented as local measurements and classical communication in quantum mechanics does. In particular, all measurements in box world are separable [26] (i.e., can be expressed as a sum of product effects), which leads to various restrictions. For instance, teleportation and entanglement swapping are impossible in box world [8,26,27]. This directly implies that entanglement swapping is also not possible in the multipartite setting (i.e., with nonwirings). A direct proof of this, which also applies to other GPTs, is obtained by following the same lines of reasoning as Lemma 2 of [50].

E. Box-world effects in other GPTs

Wirings correspond to a classical processing of inputs and outcomes from measurements and as such the inputs and outcomes of *any* GPT can be connected by wirings. We show here that from the effects derived in this work, including nonwiring effects, we can indeed construct valid effects for any other GPT.

Consider a single system that is fully characterized by n_I inputs and n_O outcomes and call the SBV effects $e_{1,1}, \dots, e_{n_I, n_O}$, where $e_{i,j}$ is the effect that has an entry 1 for the i th input and j th outcome and is zero otherwise. For N parties, the effects $\{e_{i_1, j_1} \otimes \dots \otimes e_{i_N, j_N}\}_{i_1, \dots, i_N, j_1, \dots, j_N}$ span

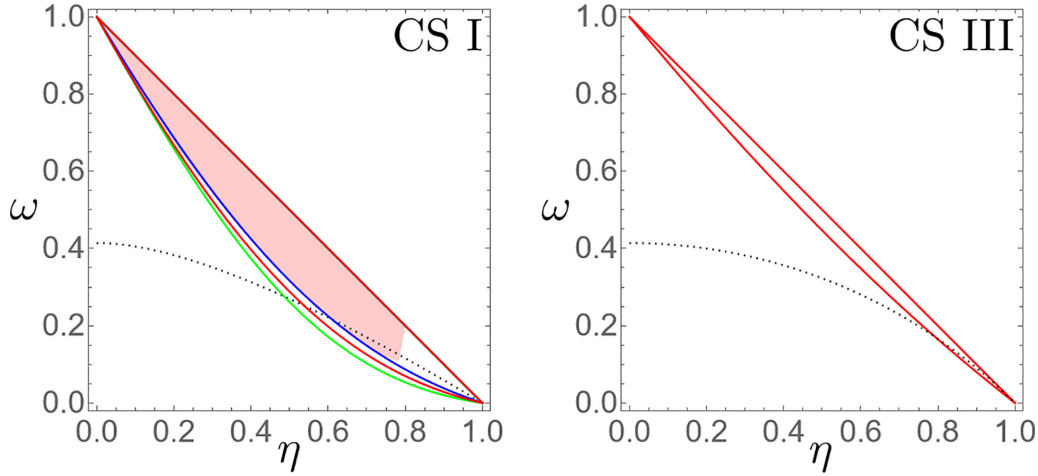


FIG. 3. The red lines show the boundary of the distillable regions using our nonwiring protocol, in different cross sections of the $(2,2,2)$ no-signaling polytope. The dotted curves represent the boundary of the quantum realizable correlations. On the left, the blue curve represents the boundary of the distillable region using the XOR protocol of [47], while the green one represents correlations that can be distilled through the OR protocols of [46]. The red shaded area shows where our nonwiring protocol achieves higher final CHSH values than the protocols of [46–49] (for the protocol of [49], the comparison has been made with both its two-copy and three-copy versions). Note that in CS III there is no two-copy protocol that can distill any of the states [46].

the effect cone of all box-world effects. Thus, the effects we derived above for N parties can be written in terms of the local effects as

$$e = \sum_{i_1, \dots, i_N=1}^{n_I} \sum_{j_1, \dots, j_N=1}^{n_O} \lambda_{i_1, \dots, i_N, j_1, \dots, j_N} e_{i_1, j_1} \otimes \dots \otimes e_{i_N, j_N},$$

where $\lambda_{i_1, \dots, i_N, j_1, \dots, j_N}$ are some coefficients (which can be taken to be positive, cf. Lemma 5). Now, as we know that any valid effect is valid on any valid state, i.e., $0 \leq \langle e, s \rangle \leq 1$, we also have

$$\begin{aligned} 0 &\leq \sum_{i_1, \dots, i_N=1}^{n_I} \sum_{j_1, \dots, j_N=1}^{n_O} \lambda_{i_1, \dots, i_N, j_1, \dots, j_N} P(j_1, \dots, j_N | i_1, \dots, i_N) \\ &\leq 1, \end{aligned}$$

where $P(j_1, \dots, j_N | i_1, \dots, i_N)$ is any no-signaling distribution. This holds because the state space in box world is in 1:1 correspondence with the set of all no-signaling distributions of which the effects $\{e_{i_1, j_1} \otimes \dots \otimes e_{i_N, j_N}\}_{i_1, \dots, i_N, j_1, \dots, j_N}$ essentially just pick out elements.

Performing local measurements on N -party states leads to nonsignaling correlations in any GPT since nonsignaling is one of the underlying assumptions. This implies that the correlations arising from performing any local measurements on an N -party system are mapped to a probability by the map defined by the coefficients $\{\lambda_{i_1, \dots, i_N, j_1, \dots, j_N}\}_{i_1, \dots, i_N, j_1, \dots, j_N}$. This means that for any GPT we can build valid effects from the ones we derived for box world, namely, using these same coefficients:

$$f = \sum_{i_1, \dots, i_N=1}^{n_I} \sum_{j_1, \dots, j_N=1}^{n_O} \lambda_{i_1, \dots, i_N, j_1, \dots, j_N} f_{i_1, j_1} \otimes \dots \otimes f_{i_N, j_N},$$

where the $f_{i,j}$ are local effects in the GPT of interest.

Note that a special case of the above is that in any GPT in which SBVs are valid local effects, all the effects found by our algorithms are directly valid. Furthermore, the state space of a GPT can always be expressed in terms of the outcome probabilities of a set of fiducial measurements. Having made the transformation needed for this representation, the SBVs are local effects.

VIII. THE SIGNIFICANCE OF WIRING AND NONWIRING OPERATIONS BEYOND GPTS

The polytope spanned by all identity effects in box world in any (N, n_I, n_O) scenario is of interest beyond the scope of box world and even beyond the scope of GPTs. This is illustrated with the following two connections to other areas of quantum foundations. This means that our Algorithm 2 may be of more general interest in the future in the sense that it allows us to construct logically consistent classical processes without causal order as well as compositions of contextuality scenarios.

A. Logically consistent classical processes

We also remark that the polytope of all identity effects in box world is equivalent to that of the logically consistent classical processes, which was computed in [23].¹¹ Since extremal effects for a single system in box world are SBVs in the $(n_I n_O)^N$ -dimensional representation (and coarse graining of these), operationally, any extremal local measurement can be performed by choosing an input and then potentially coarse

¹¹This includes not only the $\{0, 1\}$ -valued identity effects but the whole polytope.

graining the outcome, which is a classical process.¹² This explains the correspondence between performing multisystem measurements on box-world systems and the ways these classical local operations can be consistently connected. This correspondence extends: for any number of inputs and outputs the logically consistent classical processes can be understood in terms of the (identity) measurements of a GPT, namely, box world. This operational way of thinking about the logically consistent operations may aid our understanding of such operations.

B. Composing contextuality scenarios

Contextuality is a notion that captures the lack of predetermined outcomes in quantum measurements, as these may depend on the context in which the measurements are performed. There are various approaches capturing this notion, starting from the original work of Kochen and Specker [51]. In [25], the composition of contextuality scenarios was analyzed, following the approach to contextuality from [24]. According to this approach, a contextuality scenario is represented by a hypergraph (W, M) , where the vertices $e \in W$ represent events of a specific outcome occurring and hyperedges $m \in M$ correspond to collections of events that make up a measurement.¹³ These edges may overlap on some vertices, meaning that the respective event can occur as part of several measurements. In a noncontextual model, each vertex can be assigned a probability since the event occurs with the same frequency no matter which of the measurements it is part of is performed (in a contextual model the probability will in general also depend on the measurement). The probabilistic models that are allowed then further depend on the underlying theory (e.g., classical, quantum, or more general).

These contextuality scenarios can then also be considered in the multipartite regime, meaning that independent contextuality scenarios for several independent systems are turned into a single scenario for the joint system. This means that for two systems A, B for each of which a contextuality scenario $(W_A, M_A), (W_B, M_B)$ is given, a joint scenario (W_{AB}, M_{AB}) with events $e_{AB} = (e_A, e_B) \in W_{AB}, e_A \in W_A, e_B \in W_B$ is constructed in a way that the probabilistic models defined on the hypergraph satisfy the nonsignaling principle. Depending on the set M_{AB} that is constructed for this purpose, we speak about a different product. While in the case of two contextuality scenarios there is a unique way to compose them [52], namely, by means of the Foulis-Randall (FR) product, in [24,25] different ways to compose more than two such scenarios, all respecting the nonsignaling principle, were proposed.

Our multipartite $\{0, 1\}$ -valued effects can be seen as ways to combine products of deterministic local effects into multipartite measurements for our (nonsignaling) GPT systems, which indeed amounts to the same mathematical problem

as composing contextuality scenarios. Local deterministic effects correspond to the events v and the identity effects to the edges e in these multipartite scenarios. Thus, our Algorithm 1 can be seen as a way to construct compositions of contextuality scenarios. Specifically, this algorithm constructs a product known as the *disjunctive FR product* in [25]. The subset of wiring effects identified by Algorithm 3 constructs the *maximal FR product* from [25].

The existence of extremal identity effects that are not $\{0, 1\}$ valued further show that there is a more general way to combine single-system effects compatibly with nonsignaling. This reasoning could also be applied to the events in a contextuality scenario. This suggests that the hypergraph formalism for describing contextuality scenarios needs to be extended to include a new concept that generalizes the notion of a hyperedge. This generalization requires a weight for each element of the hyperedge. [This can be easily added within the matrix representation of the hypergraph which works as follows. Each column corresponds to a vertex and each row denotes a hyperedge with a 1 meaning the vertex of the corresponding column is in the hyperedge and a 0 meaning it is not. In the generalization, the matrix is no longer restricted to contain only elements of $\{0, 1\}$.] It would be interesting to explore the significance of this for noncontextuality in more detail.

IX. CONCLUSION

Characterizing measurements in theories beyond quantum theory allows us to better explore the possibilities for information processing in such theories and, in turn, helps us understand what is special about quantum theory itself. In this paper we have explored ways to generate all the effects present in a maximally nonlocal alternative theory, box world. We have been able to find all the deterministic effects in several scenarios, dividing them into wirings and nonwirings, and also found many classes of nondeterministic effect. Although we have focused on box world, theories with fewer states have fewer constraints on allowed effects (under the no-restriction hypothesis) and hence the effects we have found are applicable in a wide range of GPTs (see Sec. VIII E).¹⁴

The effects we have found are also relevant for studies of logically consistent classical processes and for compositions of contextuality scenarios, where our findings suggest the need to extend the hypergraph formalism for dealing with these. We have applied our findings to several applications, demonstrating advantages of both deterministic and nondeterministic nonwirings. In particular, we showed that, contrary to previous claims, examples of nonlocality without entanglement also appear in box world. We remark here that the *existence* of further examples of quantum nonlocality without entanglement has recently been shown in the literature, using a construction based on classical processes without causal order [54,55]. Due to the correspondences we establish in Secs. VIII A and VII E, any such example can likely be turned directly into an example for other GPTs.

In quantum theory there is often focus on entangled measurements, but there are other types of joint measurement

¹²Note that in the language of [23] a process being classical does not necessarily mean that the theory considered is classical in the GPT sense, i.e., classical probability theory.

¹³In graph theory, vertices are usually V and edges E ; the notation is different here because in a noncontextual model vertices correspond to effects and hyperedges to measurements.

¹⁴Removing states can also lead to nonseparable effects [53].

(see, e.g., [28]) whose power would be useful to explore. Since box world has no nonseparable effects and is somewhat limited with respect to its reversible dynamics [56], it would be interesting to complement our investigations with those of possible measurements in no-signaling theories with restricted nonlocality.

Aspects of this work appeared in GE's Ph.D. thesis [57].

ACKNOWLEDGMENTS

M.W. would like to thank Ä. Baumeler for discussions about logically consistent classical processes and E. Wolfe for pointing out his related work [25]. G.E. is supported by the EPSRC Grant No. EP/LO15730/1. M.W. is supported by the Lise Meitner Fellowship of the Austrian Academy of Sciences (Project No. M 3109-N).

-
- [1] S. Popescu and D. Rohrlich, Quantum nonlocality as an axiom, *Found. Phys.* **24**, 379 (1994).
- [2] L. Hardy, Quantum theory from five reasonable axioms, [arXiv:quant-ph/0101012](https://arxiv.org/abs/quant-ph/0101012).
- [3] C. A. Fuchs, Quantum mechanics as quantum information (and only a little more), [arXiv:quant-ph/0205039](https://arxiv.org/abs/quant-ph/0205039).
- [4] L. Masanes and M. P. Müller, A derivation of quantum theory from physical requirements, *New J. Phys.* **13**, 063001 (2011).
- [5] B. Dakić and C. Brukner, Quantum theory and beyond: Is entanglement special? in *Deep Beauty: Understanding the Quantum World through Mathematical Innovation*, edited by H. Halvorson (Cambridge University Press, Cambridge, 2011), p. 365–392.
- [6] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Informational derivation of quantum theory, *Phys. Rev. A* **84**, 012311 (2011).
- [7] J. S. Bell, On the Einstein-Podolsky-Rosen paradox, *Phys. Phys. Fiz.* **1**, 195 (1964).
- [8] J. Barrett, Information processing in generalized probabilistic theories, *Phys. Rev. A* **75**, 032304 (2007).
- [9] J. Barrett, L. Hardy, and A. Kent, No signalling and quantum key distribution, *Phys. Rev. Lett.* **95**, 010503 (2005).
- [10] J. Barrett, R. Colbeck, and A. Kent, Unconditionally secure device-independent quantum key distribution with only two devices, *Phys. Rev. A* **86**, 062326 (2012).
- [11] R. Colbeck and R. Renner, Free randomness can be amplified, *Nat. Phys.* **8**, 450 (2012).
- [12] R. Gallego, L. Masanes, G. De La Torre, C. Dhara, L. Aolita, and A. Acín, Full randomness from arbitrarily deterministic events, *Nat. Commun.* **4**, 2654 (2013).
- [13] W. van Dam, Implausible consequences of superstrong nonlocality, *Nat. Comput.* **12**, 9 (2013).
- [14] M. Navascués and H. Wunderlich, A glance beyond the quantum model, *Proc. R. Soc. A* **466**, 881 (2010).
- [15] M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Żukowski, Information causality as a physical principle, *Nature (London)* **461**, 1101 (2009).
- [16] T. Fritz, A. Sainz, R. Augusiak, J. B. Brask, R. Chaves, A. Leverrier, and A. Acín, Local orthogonality as a multipartite principle for quantum correlations, *Nat. Commun.* **4**, 2263 (2013).
- [17] C. Branciard, N. Gisin, and S. Pironio, Characterizing the nonlocal correlations created via entanglement swapping, *Phys. Rev. Lett.* **104**, 170401 (2010).
- [18] T. Fritz, Beyond Bell's theorem: correlation scenarios, *New J. Phys.* **14**, 103001 (2012).
- [19] J. H. Selby, A. B. Sainz, V. Magron, Ł. Czekaj, and M. Horodecki, Correlations constrained by composite measurements, *Quantum* **7**, 1080 (2023).
- [20] A. Pozas-Kerstjens, N. Gisin, and A. Tavakoli, Full network nonlocality, *Phys. Rev. Lett.* **128**, 010403 (2022).
- [21] I. Šupić, J.-D. Bancal, Y. Cai, and N. Brunner, Genuine network quantum nonlocality and self-testing, *Phys. Rev. A* **105**, 022206 (2022).
- [22] M. Weilenmann and R. Colbeck, Self-testing of physical theories, or, is quantum theory optimal with respect to some information-processing task? *Phys. Rev. Lett.* **125**, 060406 (2020).
- [23] Ä. Baumeler and S. Wolf, The space of logically consistent classical processes without causal order, *New J. Phys.* **18**, 013036 (2016).
- [24] A. Acín, T. Fritz, A. Leverrier, and A. B. Sainz, A combinatorial approach to nonlocality and contextuality, *Commun. Math. Phys.* **334**, 533 (2015).
- [25] A. B. Sainz and E. Wolfe, Multipartite composition of contextuality scenarios, *Found. Phys.* **48**, 925 (2018).
- [26] A. J. Short and J. Barrett, Strong nonlocality: A trade-off between states and measurements, *New J. Phys.* **12**, 033034 (2010).
- [27] A. J. Short, S. Popescu, and N. Gisin, Entanglement swapping for generalized nonlocal correlations, *Phys. Rev. A* **73**, 012101 (2006).
- [28] C. H. Bennett, D. P. DiVincenzo, C. A. Fuchs, T. Mor, E. Rains, P. W. Shor, J. A. Smolin, and W. K. Wootters, Quantum nonlocality without entanglement, *Phys. Rev. A* **59**, 1070 (1999).
- [29] E. Chitambar, D. Leung, L. Mančinska, M. Ozols, and A. Winter, Everything you always wanted to know about LOCC (but were afraid to ask), *Commun. Math. Phys.* **328**, 303 (2014).
- [30] A. Chefles, Condition for unambiguous state discrimination using local operations and classical communication, *Phys. Rev. A* **69**, 050307(R) (2004).
- [31] E. Chitambar, W. Cui, and H.-K. Lo, Increasing entanglement monotonies by separable operations, *Phys. Rev. Lett.* **108**, 240504 (2012).
- [32] A. M. Childs, D. Leung, L. Mančinska, and M. Ozols, A framework for bounding nonlocality of state discrimination, *Commun. Math. Phys.* **323**, 1121 (2013).
- [33] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Probabilistic theories with purification, *Phys. Rev. A* **81**, 062348 (2010).
- [34] L. Hardy, Foliabile operational structures for general probabilistic theories, in *Deep Beauty: Understanding the Quantum World through Mathematical Innovation*, edited by H. Halvorson (Cambridge University Press, Cambridge, 2011), p. 409–442.
- [35] B. S. Tsirelson, Some results and problems on quantum Bell-type inequalities, *Hadronic J. Suppl.* **8**, 329 (1993).

- [36] D. Rosset, J.-D. Bancal, and N. Gisin, Classifying 50 years of Bell inequalities, *J. Phys. A: Math. Theor.* **47**, 424022 (2014).
- [37] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevA.108.062212> for data files containing the effects found.
- [38] T. Christof and A. Loebel, POlyhedron Representation Transformation Algorithm (PORTA).
- [39] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic, London, 1976).
- [40] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).
- [41] A. J. Short and S. Wehner, Entropy in general physical theories, *New J. Phys.* **12**, 033023 (2010).
- [42] J. Löfberg, YALMIP : A toolbox for modeling and optimization in MATLAB, in *IEEE International Conference on Robotics and Automation (IEEE Cat. No. 04CH37508), Taipei, Taiwan* (IEEE, Piscataway, NJ, 2004).
- [43] The MOSEK optimization toolbox for MATLAB manual, Version 9.0.105, <http://docs.mosek.com/9.0/toolbox/index.html>.
- [44] S. S. Bhattacharya, S. Saha, T. Guha, and M. Banik, Nonlocality without entanglement: Quantum theory and beyond, *Phys. Rev. Res.* **2**, 012068(R) (2020).
- [45] B. Groisman and L. Vaidman, Nonlocal variables with product-state eigenstates, *J. Phys. A: Math. Gen.* **34**, 6881 (2001).
- [46] G. Eftaxias, M. Weilenmann, and R. Colbeck, Advantages of multi-copy nonlocality distillation and its application to minimizing communication complexity, *Phys. Rev. Lett.* **130**, 100201 (2023).
- [47] J. Allcock, N. Brunner, N. Linden, S. Popescu, P. Skrzypczyk, and T. Vértesi, Closed sets of nonlocal correlations, *Phys. Rev. A* **80**, 062107 (2009).
- [48] P. Høyer and J. Rashid, Optimal protocols for nonlocality distillation, *Phys. Rev. A* **82**, 042118 (2010).
- [49] M. Forster, S. Winkler, and S. Wolf, Distilling nonlocality, *Phys. Rev. Lett.* **102**, 120401 (2009).
- [50] M. Weilenmann and R. Colbeck, Toward correlation self-testing of quantum theory in the adaptive Clauser-Horne-Shimony-Holt game, *Phys. Rev. A* **102**, 022203 (2020).
- [51] S. Kochen and E. P. Specker, The problem of hidden variables in quantum mechanics, *J. Math. Mech.* **17**, 59 (1967).
- [52] D. J. Foulis and C. H. Randall, Empirical logic and tensor products, in *Interpretations and Foundations of Quantum Theory* (B.I.-Wissenschaftsverlag, Zurich, 1981).
- [53] P. Skrzypczyk and N. Brunner, Couplers for non-locality swapping, *New J. Phys.* **11**, 073014 (2009).
- [54] R. Kunjwal and Ämin Baumeler, Trading causal order for locality, *Phys. Rev. Lett.* **131**, 120201 (2023).
- [55] S. Akibue, M. Owari, G. Kato, and M. Murao, Entanglement-assisted classical communication can simulate classical communication without causal order, *Phys. Rev. A* **96**, 062331 (2017).
- [56] D. Gross, M. Müller, R. Colbeck, and O. C. O. Dahlsten, All reversible dynamics in maximally nonlocal theories are trivial, *Phys. Rev. Lett.* **104**, 080402 (2010).
- [57] G. Eftaxias, Theory-independent topics towards quantum mechanics: ψ -ontology and nonlocality distillation, Ph.D. thesis, Quantum Engineering Centre for Doctoral Training, University of Bristol, UK, 2022.