The space of logically consistent classical processes without causal order

Ámin Baumeler$^{1,2}$ and Stefan Wolf$^{1,2}$

$^1$ Faculty of Informatics, Università della Svizzera italiana, Via G. Buffi 13, 6900 Lugano, Switzerland
$^2$ Facoltà indipendente di Gandria, Lunga scala, 6978 Gandria, Switzerland

E-mail: baumea@usi.ch

Keywords: correlations, causality, no causal order, games, polytope

Abstract

Classical correlations without predefined causal order arise from processes where parties manipulate random variables, and where the order of these interactions is not predefined. No assumption on the causal order of the parties is made, but the processes are restricted to be logically consistent under any choice of the parties’ operations. It is known that for three parties or more, this set of processes is larger than the set of processes achievable in a predefined ordering of the parties. Here, we model all classical processes without predefined causal order geometrically and find that the set of such processes forms a polytope. Additionally, we model a smaller polytope—the deterministic-extrema polytope—where all extremal points represent deterministic processes. This polytope excludes probabilistic processes that must be—quite unnaturally—fine-tuned, because any variation of the weights in a decomposition into deterministic processes leads to a logical inconsistency.

1. Motivation and main result

An assumption often made in physical theories, sometimes implicitly, is the existence of a global time. In particular, quantum theory is formulated with time as an intrinsic parameter. If one relaxes this assumption by requiring local validity of some theory and logical consistency only, then a larger set of correlations can be obtained, called correlations without predefined causal order. The processes that lead to such correlations are called processes without predefined causal order. Two motivations to study such correlations are quantum gravity and quantum non-locality. Quantum gravity motivates this research in the sense that on the one hand, relativity is a deterministic theory equipped with a dynamic spacetime; on the other hand, quantum theory is a probabilistic theory embedded in a fixed spacetime. This suggests that quantum gravity is relaxed in both aspects, i.e., it is a probabilistic theory equipped with a dynamic spacetime [1]. Quantum non-local correlations [2–4] motivate this study since the possibility of a satisfactory causal explanation [5] for such correlations is questionable [3, 6–13]. Dropping the notion of a global time or of an a priori spacetime—as has been suggested from different fields of research [14–23]—dissolves this paradox. This can be achieved by defining causal relations based on free randomness (see figure 1) as opposed to defining free randomness based on causal relations [24, 25]. Such an approach gives a dynamic character to causality; causal connections are not predefined but are derived from the observed correlations.

Relaxations of quantum theory where the assumption of a global time is dropped have recently been studied widely [1, 26–45] (see [46] for a review). Our work follows the spirit of an operational quantum framework for such correlations developed by Oreshkov et al [31]. Some correlations appearing in their quantum framework—for two parties or more—cannot be simulated by assuming a predefined causal order of the parties. Such correlations are termed non-causal. Analogously to non-locality, non-causal correlations could be witnessed by violating so-called causal inequalities [31, 32, 35, 43]. All causal inequalities in the two-party scenario and for binary inputs and outputs are presented in [43]. In a previous work [35], we showed that in the classical limit of the quantum framework, i.e., if it is restricted to probability theory, classical non-causal correlations can arise as well. This result holds for three parties or more. In the present work we follow this path and give a representation
of all classical—as opposed to quantum—processes without predefined causal order as polytopes. Such a representation helps in optimizing winning strategies for causal games [31, 43]—the optimization problem can be stated as a linear program—and for finding new causal games.

First, we present the framework of classical correlations without predefined causal order. Then, we describe the polytope of processes that lead to such correlations implicitly and explicitly for scenarios with up to three parties and binary inputs and outputs. In the general case, we give an implicit description of the polytope. In addition, we construct the smaller polytope of classical processes without predefined causal order where all extremal points describe deterministic processes. We call this polytope the deterministic-extrema polytope. The processes from this polytope can be thought of as being ‘more physical’ in the sense that its extremal points are not proper mixtures of logically inconsistent processes [47], i.e., this set contains processes that can be written as a convex combination of deterministic ones from within the polytope only. Our motivation for this is that some proper mixtures need to be fine-tuned [13], i.e., tiny variations of the mixtures renders the processes logically inconsistent. The fine-tuned proper mixtures are the probabilistic extremal points of the larger polytope. A qualitative representation of these polytopes is given in figure 2.

2. Modelling classical correlations without predefined causal order

2.1. Causality, predefined causal order, and a framework of classical correlations without predefined causal order

We describe an operational framework without global assumptions (other than logical consistency). Causal relations are defined as in the interventionists’ approach to causality [48, 49]: Outputs can be correlated to inputs and inputs are manipulated freely (see figure 1). Defining causality based on free randomness is the converse approach to the one used in recent literature [24, 25]; there, free randomness is defined based on causal relations.

**Definition 1.** (Causality [35]). For two correlated random variables $X$ and $A$, where $X$ is an output and $A$ is an input, i.e., $A$ is chosen freely, we say that $X$ is in the causal future of $A$, or equivalently, that $A$ is in the causal past of $X$, denoted by $X \triangleright A$ or $A \triangleleft X$. The negations of these relations are denoted by $\nLeftarrow$ and $\nRightarrow$.

Consider $N$ parties $\{S_i\}_{0 \leq i < N}$, where party $S_i$ has access to an input random variable $A_i$ and generates an output random variable $X_i$. This allows us to causally order parties: if $A_i$ is correlated to $X_{i+1}$, then $S_i$ is in the causal past of $S_{i+1}$ ($S_i \nLeftarrow S_{i+1}$). To simplify the presentation, we write $\vec{X} = (X_0, ..., X_{N-1})$ and likewise for $\vec{A}$, $\vec{O}$, and $\vec{I}$.
Figure 3. A single party $S_j$ describes a stochastic process $P_{X_j,O_j|I_j}$. The variables $A_j$ and $X_j$ model the input and the output. The variable $I_j$ is obtained from the environment $E$; the party $S_j$ feeds the variable $O_j$ into the same environment.

**Definition 2.** (Two-party predefined causal order). A two-party predefined causal order is a causal ordering of party $S$ with input $A$, output $X$, and party $T$ with input $B$, output $Y$, such that the distribution $P_{X,Y|A,B}$ can be written as a convex combination of one-way signaling distributions

$$P_{X,Y|A,B} = pP_{X|A,B} + (1-p)P_{Y|A,B,X},$$

for some $0 \leq p \leq 1$.

A definition for multi-party predefined causal order is given in [42]. Such a definition turns out to be more subtle since a party $S_j$ in the causal past of some other parties $\{S_k\}_k$ can in principle influence everything in her causal future; in particular, $S_j$ can influence the causal order of the parties $\{S_k\}_k$. We just state a lemma that follows from such a definition and that is sufficient to prove our claims.

**Lemma 1.** (Necessary condition for predefined causal order). A necessary condition for a predefined causal order is that the probability distribution $P_{X|I}$ can be written as a convex combination

$$P_{X|I} = \sum_k p_k P_k,$$

with $\sum_k p_k = 1$ and $\forall k : p_k \geq 0$, such that in every distribution $P_k$ at least one party is not in the causal future of any other party, i.e.

$$\forall k \exists i \forall j \not\exists i : S_j \not\subseteq S_i,$$

where $\subseteq$ stands for the causal relation that is deduced from the distribution $P_k$.

In the framework without predefined causal order, each party $S_j$ receives a random variable $I_j$ from the environment $E$ on which $S_j$ can act. After the interaction with $I_j$, party $S_j$ outputs a random variable $O_j$ to the environment. Both random variables $I_j$ and $O_j$ are output random variables. The only input random variable a party has is $A_j$. The operation of $S_j$ is a stochastic process mapping $A_j$, $I_j$ to $X_j$, $O_j$ (see figure 3). A stochastic process is a probability distribution over the range conditioned on the domain; in this case, the stochastic process of party $S_j$ (which in the following will also be called the local operation of party $S_j$) is $P_{X_j,O_j|A_j,I_j}$. All parties are allowed to apply any possible operation described by probability theory. Furthermore, they are isolated from each other, which means that they can interact only through the environment. Because we do not make global assumptions (beyond logical consistency), the most general picture is that the random variables that are sent from the environment $E$ to the parties are the result of a map on the random variables fed back by all parties to the same environment $E$ (see figure 4). Such a composition of parties with the environment combines states and communication channels in one framework.

A party $S_j$ has access to the four random variables $X_j$, $O_j$, $I_j$, and $A_j$ where $A_j$ is chosen freely. If we consider all parties together, we should get a probability distribution $P_{X_{\ell},O_{\ell}|A_{\ell},I_{\ell}}$. Furthermore, we ask the environment $E$ to be a multi-linear functional of all local operations. The motivation for this is that linear combinations of local operations should carry through to the probabilities $P_{X_{\ell},O_{\ell}|A_{\ell},I_{\ell}}$. This brings us to a definition of logical consistency.

**Definition 3.** (Logical consistency). An environment $E$ is called logically consistent if and only if it is a multi-linear positive map on any choice of local operations $\{P_{X_j,O_j|A_j,I_j}\}_{0 \leq j < N}$ of all parties such that the composition of $E$ with the local operations results in a probability distribution $P_{X_{\ell},O_{\ell}|A_{\ell},I_{\ell}}$. 

---

**References:**

[42] A. Baumeler and S. Wolf, *New J. Phys.*, 18(2016) 013036.
The linearity and positivity conditions from definition 3 imply theorem 1, which states that the environment must be a stochastic process (conditional probability distribution).

**Theorem 1.** (Logical consistent environment as stochastic process). The environment $E$ is a stochastic process $P_{\bar{O}|\bar{I}}$ that maps $\bar{O}$ to $\bar{I}$.

**Proof.** The environment is a multi-linear positive map $E$ on the probabilities (we omit the arguments for the sake of presentation)

$$p_j := P_{X_j, O_j|A_j, j}(x_j, o_j, a_j, i_j)$$

that party $S_j$ outputs $o_j$ to the environment and generates $x_j$ conditioned on the setting $a_j$ and on $I_j = i_j$.

Therefore, we write

$$P_{\bar{X}|\bar{O}, \bar{I}}(\bar{x}, \bar{o}, \bar{a}) = E(p_0 \cdots p_{N-1}).$$

Since $E$ is a multi-linear positive map and since it depends on $\bar{O}$ and $\bar{I}$ only, the above probability can be written as

$$P_{\bar{X}|\bar{O}, \bar{I}}(\bar{x}, \bar{o}, \bar{a}) = E(\bar{o}, \bar{t})p_0 \cdots p_{N-1},$$

(1)

where $E(\bar{o}, \bar{t})$ is a number. This number must be non-negative, as otherwise the above expression (1) is not a probability. By fixing $\bar{A} = \bar{a}$ and by summing over $\bar{x}$, we get

$$P_{\bar{I}|\bar{o}|\bar{A} = \bar{a}}(\bar{t}, \bar{o}) = \sum_{\bar{x}} E(\bar{o}, \bar{t}) = E(\bar{o}, \bar{t})\sum_{\bar{x}} p_0 \cdots p_{N-1} = E(\bar{o}, \bar{t})p_0' \cdots p'_{N-1},$$

where

$$p'_j := P_{O_j|I_j, A_j = a_j}(o_j, i_j).$$

Let us fix the local operations $p'_j$ of all parties to be

$$P_{O_j|I_j, A_j = a_j}(o_j, i_j) = \begin{cases} 1 & o_j = 0, \\ 0 & \text{otherwise}. \end{cases}$$

From the total-probability condition we obtain

$$\sum_{\bar{o}, \bar{a}} E(\bar{o}, \bar{t}) p_0' \cdots p'_{N-1} = \sum_{\bar{i}} E(\bar{o}, \bar{t}) = 1.$$

By repeating this calculation for different choices of local operations where the parties deterministically output a value, we get

$$\forall \bar{o} : \sum_{\bar{i}} E(\bar{o}, \bar{t}) = 1.$$

Therefore, $E$ is a stochastic process $P_{\bar{I}|\bar{O}}$. \qed

Figure 4. The box $E$ describes the environment. Because no predefined causal order is assumed between the parties, the random variable obtained by the parties is the result of $E$ applied to the outgoing random variable of all parties. This picture combines states and channels, i.e., signaling and no-signaling correlations. For example, assume that $S_0$ is in the causal past of all other parties. In that case, the random variable $I_0$ is constant, whereas the random variable $I_j$ could depend on $A_j$. For three parties or more, this framework gives rise to a new quality: $E$ can describe a map where no $I_j$ is a constant, yet where no contradiction arises. Such correlations are called non-causal. Similarly to the parties, the box $E$ is a stochastic process $P_{O_0, \ldots, O_{N-1}|I_0, \ldots, I_{N-1}}$. 

New J. Phys. 18 (2016) 013036 A Baumeler and S Wolf
The following corollary follows from theorem 1.

**Corollary 1.** A logical consistent environment $P_{I|O}$ fulfills the property that under any choice of the local operations $\{P_{X,O}|A_i\}_{0 \leq i < N}$ of all parties, the expression $P_{I|O} \prod_{j=0}^{N-1} P_{X_j,O_j|A_i,j}$ form a conditional probability distribution $P_{X,I,O\mid A}$.

Note that not every conditional distribution $P_{I|O}$ is logically consistent. Some stochastic processes lead to *grandfather-paradox-type* inconsistencies. Consider the following two extreme examples of such inconsistencies. We describe the examples in the single-party scenario as depicted in figure 5 and where $O$, $I$, $X$, and $A$ are binary random variables.

**Example 1.** Let the environment as well as the party $S$ forward the random variable, i.e., the operation of the environment is

$$P_{I|O}(i, o) = \begin{cases} 1 & \text{if } i = o, \\ 0 & \text{otherwise,} \end{cases}$$

and the operation of the party $S$ is

$$P_{X,O|A,i}(x, o, a, i) = \begin{cases} 1 & \text{if } i = x, \\ 0 & \text{otherwise.} \end{cases}$$

Since the environment $E$ and the party $S$ forward the random variable, we have $\Pr(O = 1) = 1$. However, it is unclear what value the probability $P_{I}(0)$ should take. This is also known as the *causal-loop paradox*.

**Example 2.** We alter the local operation of party $S$ to *negate* the binary random variable

$$P_{X,O|A,i}(x, o, a, i) = \begin{cases} 1 & \text{if } i \oplus 1 = x, \\ 0 & \text{otherwise.} \end{cases}$$

Now, we are faced with the *grandfather paradox*: if party $S$ receives $i = 1$ from the environment, then she sends the value $o = 0$ to the environment. But in that case, she should receive $i = 0$ and not $i = 1$.

### 2.2. Mathematical model of states, operations, evolution, and composition

Let $\{q_0, q_1, \ldots\}$ be the sample space of a random variable $Q$ with the probability measure $P_Q$.

**Definition 4.** (States, operations, evolution, and composition). We represent a state corresponding to a random variable $P_Q$ as the probability vector

$$\bar{P}_Q = (P_Q(q_0), P_Q(q_1), \ldots)^T.$$  

A stochastic process $P_{R|Q}$ from $Q$ to a random variable $R$ with sample space $\{r_0, r_1, \ldots\}$ describes an operation and is modeled by the stochastic matrix

$$\hat{P}_{R|Q} = \begin{pmatrix} P_{R|Q}(r_0, q_0) & P_{R|Q}(r_0, q_1) & \cdots \\ P_{R|Q}(r_1, q_0) & P_{R|Q}(r_1, q_1) & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}.$$

The result $P_R$ of evolving the random variable $P_Q$ through the operation $P_{R|Q}$ is given by the matrix multiplication

$$\bar{P}_R = \hat{P}_{R|Q} \bar{P}_Q.$$

Finally, vectors and matrices are composed in parallel using the Kronecker product $\otimes$.  

---

![Figure 5](image.png)

*Figure 5.* Party $S$ is described by $P_{X,O|A,i}$ and the environment $E$ is $P_{I|O}$. 

---

**Remark:** The diagram illustrates the flow of information in a single-party scenario. The environment $E$ and the party $S$ forward the random variables. The joint probability distribution $P_{X,I,O\mid A}$ is represented by the Kronecker product $\otimes$.
For example, by this definition, the output of a stochastic process \( P_{R|Q_0,Q_1} \) taking two inputs and producing one output is expressed by

\[
P_{R|Q_0,Q_1}(\tilde{P}_{Q_0} \otimes \tilde{P}_{Q_1}).
\]

2.3. Set of logically consistent processes without predefined causal order

We derive the conditions on the environment \( E \) (stochastic process) such that it is logically consistent. For simplicity, we start with the single-party scenario as depicted in figure 5; the party is denoted by \( S \) and the environment by \( E \). We can further simplify our picture by fixing the value of \( A \) to \( a \) and by summing over \( X \):

\[
P_{P|X=x,\tilde{A}=a,\widetilde{X}}(\mathcal{A}) = \sum_{x} P_{X=x,\tilde{A}=a,\widetilde{X}} = P_{P|\tilde{A}=a}.
\]

The stochastic process of the environment \( E \) is \( P_{IO} \). For now, let us assume that \( S \) performs a deterministic operation \( D_{Oi} \). This assumption is dropped later. The operation applied by \( S \) can be written as a function \( i \rightarrow f(i) \), where \( i \) is a deterministic input value. By embedding \( f \) into the process of \( E \), we get

\[
P_{P|i=I}(f(i)).
\]

This can be interpreted as a probability measure of party \( S \) receiving the value \( i \) from the environment \( E \):

\[
Q(i) = P_{P|i}(f(i)).
\]

For \( Q(i) \) to represent a probability measure, the values of \( Q \) for every deterministic value \( i \) must be non-negative and have to sum up to 1:

\[
\forall i: Q(i) \geq 0, \quad \sum_i Q(i) = 1.
\]

We express both conditions in the matrix picture. Non-negativity is achieved whenever all entries of the matrix \( \tilde{P}_{I|O} \) are non-negative. The total-probability conditions are formulated in the following way. The value \( f(i) \) that is fed into the environment \( E \) is

\[
\tilde{D}_{O|I=I} = \tilde{D}_{O|I} \cdot I.
\]

The matrix \( \tilde{P}_{I|O} \) fixed to providing the state \( \tilde{I} \) to the party \( S \) is

\[
\tilde{P}_{I=I|O} = \tilde{I} \cdot \tilde{P}_{I|O}.
\]

Therefore, the probability of party \( S \) observing \( i \) is

\[
Q(i) = \tilde{I} \cdot \tilde{P}_{I|O} \tilde{D}_{O|I} \tilde{I},
\]

and the law of total probability requires

\[
\text{Tr}(\tilde{P}_{I|O} \tilde{D}_{O|I}) = 1.
\]

This condition remains the same if we relax the input to a stochastic input and the operation of \( S \) to a stochastic process \( P_{Qi} \). The reason for this is that any stochastic input can be written as a convex combination of deterministic inputs, and any stochastic process can be written as a convex combination of deterministic operations. Therefore, the logical-consistency requirement asks the environment \( E \) to be restricted to those processes \( \tilde{E} \) where, under any choice of the local operation \( P_{Qi} \) of party \( S \), the law of total probability

\[
\text{Tr}(\tilde{E}_{O|I}) = 1
\]

and the non-negativity condition

\[
\forall i, j: \tilde{E}_{i,j} \geq 0
\]

hold. Because a stochastic process can be written as a convex mixture of deterministic operations, it is sufficient to ask for

\[
\forall D \in \mathcal{D} : \text{Tr}(\tilde{E}D) = 1
\]

\[
\forall i, j: \tilde{E}_{i,j} \geq 0
\]

for every operation \( \tilde{D} \) from the set \( \mathcal{D} \) of all deterministic operations. Thanks to linearity, we can straightforwardly extend these requirements to multiple parties, and arrive at theorems 2 and 3.

**Theorem 2.** *Total probability.* The law that the sum of the probabilities over the exclusive states the parties receive is 1 is satisfied if and only if
The reverse direction of the bijection follows from the description above.

Analogously, the stochastic matrix

\[ \sum_{k} |k\rangle_{H_0} \otimes d(\hat{E}|k\rangle) \sum_{\ell} \langle \ell|_{H_1} \right] 

where \( \hat{E} \) represents a deterministic operation of party \( S_j \).

**Theorem 3.** (Non-negative probabilities). The law that the probability of the parties observing a state is non-negative is satisfied if and only if

\[ \forall i, j : \hat{E}_{ij} \geq 0. \]

### 2.4. Equivalence to the quantum correlations framework in the classical limit

The ingredients of the framework by Oreshkov et al. [31] are process matrices and local operations—described by matrices as well. All the matrices are completely positive trace-preserving quantum maps in the Choi–Jamiołkowski [31, 32] picture. In the classical limit, the matrices become diagonal in the computational basis [31, 35]. In the single-party scenario, the process matrix \( W \) is a map from the Hilbert space \( \mathcal{H}_0 \) to the Hilbert space \( \mathcal{H}_1 \). The party’s local operation \( A \) then again is a map from the Hilbert space \( \mathcal{H}_1 \) to the Hilbert space \( \mathcal{H}_0 \). The conditions a process matrix \( W \) in a single-party scenario has to fulfill [31] are

\[ \forall A \in \mathcal{M} : \text{Tr}(WA) = 1, \]

\[ W \geq 0, \]

where \( \mathcal{M} \) is the set of all completely positive trace-preserving maps from the space \( \mathcal{H}_1 \) to the space \( \mathcal{H}_0 \). Intuitively, the condition given by equation (4) ‘short-circuits’ both maps and enforces the probabilities of the outcomes to sum up to 1.

**Theorem 4.** (Equivalence). The quantum framework given by equations (4) and (5) in the classical limit is equivalent to the description of classical correlations without predefined causal order given by equations (2) and (3).

**Proof.** The process matrix \( W \) in the quantum framework corresponds to the stochastic process of the environment \( \hat{E} \) in our framework, and the local operations correspond to the stochastic process of the parties. We show a bijection between process matrices and stochastic processes of the environment, and between local operations and stochastic processes of the parties.

A stochastic matrix \( \hat{E} \), representing the environment \( E \) in our framework, can be translated into the quantum framework by

\[ W_{\hat{E}} = \sum_{k} |k\rangle_{H_0} \otimes d(\hat{E}|k\rangle) \sum_{\ell} \langle \ell|_{H_1} \right] 

where \( |k\rangle \) and \( |\ell\rangle \) are computational-basis states of the same dimension as \( \hat{E} \), and where the subscripts denote the respective Hilbert spaces. This completely positive trace-preserving map (expressed in the Choi–Jamiołkowski picture) acts in the same way as the stochastic matrix \( \hat{E} \): the state \( |k\rangle \) is mapped to \( \hat{E}|k\rangle \). The function \( d(\rho) \) takes the matrix \( \rho \) and cancels all off-diagonal terms, i.e.

\[ d(\rho) = \sum_{m} |m\rangle \langle m| \rho |m\rangle \langle m|. \]

We can rewrite \( W_{\hat{E}} \) as

\[ W_{\hat{E}} = \sum_{k} |k\rangle_{H_0} \otimes \sum_{m} |m\rangle \langle m| \hat{E} |k\rangle \langle m|_{H_1}. \]

Analogously, the stochastic matrix \( \hat{P}_{O|F} \) of the party can be translated into the quantum framework and becomes

\[ A_{\hat{P}_{O|F}} = \sum_{k', m'} |m'\rangle \langle m'| \hat{P}_{O|F} |k'\rangle \langle m'|_{H_0} \otimes |k'\rangle \langle k'|_{H_1}. \]

The reverse direction of the bijection follows from the description above.

Now, we show that the conditions (4) and (5) in a single-party scenario on a process matrix \( W \) coincide with the conditions (2) and (3) in our framework. The non-negativity condition (5) forces the probabilities of the outputs of \( W \) to be non-negative; the same holds for the condition (3) in our framework. That the condition (4) coincides with the condition (2) is shown below. Forcing \( W \) and \( A \) to be diagonal in the computational basis gives
Substituting $W$ with $W_E$ and $A$ with $A_{POI}$ yields

$$\sum_{i,j} \sum_{i,j} \langle i|j\rangle \langle k|i\rangle \langle j|m\rangle \langle m|k\rangle = \sum_{i,j} \sum_{i,j} \langle i|j\rangle \langle k|i\rangle \langle j|m\rangle \langle m'|k'\rangle \langle k'|j\rangle = \sum_{i,j} \sum_{i,j} \langle i|j\rangle \langle k|i\rangle \langle j|m\rangle \langle m'|k'\rangle \langle k'|j\rangle = \text{Tr}(\hat{E}_{POI}),$$

which proves the claim. The multi-party case follows through linearity.

3. Polytope of classical processes without predefined causal order

3.1. Polytopes
Convex polytopes can be represented in two different ways: the $H$-representation is a list of half-spaces where the intersection is the polytope, and the $V$-representation is a list of the extremal points of the polytope. Algorithms like the double-description method [53, 54] enumerate all extremal points of the polytope given the $H$-representation. We used cdd+ [55] for vertex enumeration. The inverse problem is solved by its dual: a convex-hull algorithm.

Here, we derive the polytope of classical processes without predefined causal order. This polytope is represented by the dashed–dotted lines in figure 2. A projection of the polytope for three parties and binary inputs/outputs onto a plane is given in figure 6.

3.2. Single party, binary input, and binary output
We start with the polytope for one party (see figure 5) with a binary input and a binary output. In this case, a process is described by a square matrix of dimension 2. The most general process of the environment $E$ is

$$E = P_{11}O = \begin{pmatrix} w_0 & w_1 \\ w_2 & w_3 \end{pmatrix},$$

consisting of four variables. The deterministic operations party $S$ can apply are

$$D_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix},$$

$$D_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad D_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$
where $\hat{D}_0$, $\hat{D}_1$ produce a constant 0, 1, respectively, and where the matrix $\hat{D}_2$ is the identity and $\hat{D}_3$ the negation. The equalities

$$\text{Tr}(\hat{E}\hat{D}_0) = 1, \quad (6)$$

$$\text{Tr}(\hat{E}\hat{D}_1) = 1, \quad (7)$$

$$\text{Tr}(\hat{E}\hat{D}_2) = 1, \quad (8)$$

enforce

$$\text{Tr}(\hat{E}\hat{D}_3) = 1.$$

This is shown as follows:

$$\text{Tr}(\hat{E}\hat{D}_0) + \text{Tr}(\hat{E}\hat{D}_1) + \text{Tr}(\hat{E}\hat{D}_2) = (w_0 + w_2) + (w_1 + w_3) + (w_0 + w_3)$$

$$= 2(w_0 + w_3) + w_1 + w_2$$

$$= 2\text{Tr}(\hat{E}\hat{D}_2) + \text{Tr}(\hat{E}\hat{D}_3).$$

By eliminating three variables using the total-probability conditions (6)–(8) from above, we get

$$\hat{P}_{IO} = \begin{pmatrix} w_0 & w_0 \\ 1 - w_0 & 1 - w_0 \end{pmatrix}$$

with the non-negativity conditions

$$w_0 \geq 0,$$

$$1 - w_0 \geq 0.$$

This solution set is a one-dimensional polytope with the extremal points 0 and 1. All solutions describe a state. This implies that all correlations that can be obtained in this framework with a single party and binary input and output, can also be obtained in a framework without feedback, i.e., these correlations can be obtained causally (see figure 7).

### 3.3. Two parties, binary inputs, and binary outputs

In the two-party case with a binary input and a binary output for each party, the process $\hat{E} = \hat{P}_{IO|O_0}$ of the environment is described by a square matrix of dimension $2^2$. The conditions are

$$\forall i, j \in \{0, 1, 2\} : \text{Tr}(\hat{E}(\hat{D}_i \otimes \hat{D}_j)) = 1,$$

$$\forall i, j : \hat{E}_{i,j} \geq 0. \quad (9)$$

With a similar argument as above, one can show that the operation $\hat{D}_3$ does not need to be considered for either party. The matrix $\hat{E}$ consists of $4^2$ unknowns, out of which $3^2$ are eliminated by the total-probability conditions given by equation (9). Thus, we are left with seven unknowns, forming a seven-dimensional polytope with 16 inequalities.

The resulting $V$-representation of the polytope consists of 12 extremal points, all of which represent deterministic processes:

---

**Figure 7.** All logically consistent single-party correlations that can be obtained with a feedback channel (see figure 5) can be simulated without a feedback channel.

---
In the following, we use \( A = S_0 \) and \( B = S_1 \). The first four processes \( \tilde{E}_0, \tilde{E}_1, \tilde{E}_2, \tilde{E}_3 \) represent the four constants \((0, 0), (0, 1), (1, 0), (1, 1)\) as inputs to the parties \( A \) and \( B \). The next four processes represent a constant input to party \( A \) (processes \( \tilde{E}_4 \) and \( \tilde{E}_5 \) produce the constant 0, and the other two processes produce the constant 1) and a channel from party \( A \) to party \( B \); the processes \( \tilde{E}_4 \) and \( \tilde{E}_6 \) describe the identity channel, and \( \tilde{E}_5 \) and \( \tilde{E}_7 \) describe the bit-flip channel. The last four processes are analogous, with a channel from \( B \) to \( A \) and where party \( B \) receives a constant. All these 12 processes act deterministically on bits for two parties where at least one party receives a constant (see figure 8). Therefore, every such channel can be simulated in a causal fashion. This result generalized to higher dimensions was already shown by taking the classical limit of the framework for quantum correlations without predefined causal order [31].

3.4. Three parties, binary inputs, and binary outputs

The process of the environment \( E \) in a three party setup with binary inputs and outputs is described by a square matrix \( \tilde{E} = \tilde{P}_{D_D',K,O,O'} \) of dimension \( 2^3 \). The matrix \( \tilde{E} \) consists of \( 4^3 \) variables, out of which \( 3^3 \) can be eliminated with the total-probability conditions

\[
\forall i, j, k \in \{0, 1, 2\} : \text{Tr}(\tilde{E} (\tilde{D}_i \otimes \tilde{D}_j \otimes \tilde{D}_k)) = 1,
\]

Figure 8. (a) Both parties \( A \) and \( B \) receive a constant each. (b) Party \( A \) receives a constant and sends a bit through the identity \( (c = 0) \) or the bit-flip \( (c = 1) \) channel to \( B \). (c) Same as (b), where the parties are interchanged.
resulting in a 37-dimensional polytope with 43 linear constraints (non-negative probabilities):

\[ \forall i, j : \hat{E}_{ij} \geq 0. \] (11)

Solving this polytope yields 710'760 extremal points. Only 744 extremal points out of these 710'760 are deterministic, i.e., consist of 0–1 values; the remaining extremal points are so-called proper mixtures of logically inconsistent processes. Such proper mixtures are not convex combinations of deterministic extremal points inside the polytope, but are convex combinations of deterministic points where some lie outside of the polytope—any process from outside of the polytope leads to logical inconsistencies. Interestingly, this smaller polytope (hence, also the polytope described by the equations (10) and (11)) consists of processes that cannot be simulated using a predefined causal order, i.e., processes where no party receives a constant, implying that every party causally succeeds some other party. The 744 deterministic extremal points are discussed in section 4 along with the general polytope restricted to the deterministic extremal points.

3.5. General case

We describe the polytope for logically consistent classical processes without predefined causal order in the general case. Let \( n \) be the number of parties and let \( d \) be the dimension of the states entering and leaving every laboratory. This leaves us with a \( d^n \times d^n \) stochastic matrix \( \hat{E} \) describing the environment. Every party can perform an operation that is a convex mixture of all \( d \times d \) deterministic operations. The set of all deterministic operations is denoted by \( \Delta \). For every party, under any choice of deterministic operation \( D \in \Delta \), the trace of the environment \( \hat{E} \) multiplied with the local operations is constrained to give 1 (see theorem 2). However—as in the binary-input/output case above—, some of these constraints are redundant.

**Theorem 5.** *(Sufficient set for total-probability conditions).* The total-probability conditions to this family of operations

\[
\hat{D}_{ij} = \begin{pmatrix}
1 & 1 & 1 & \ldots & 1 & 0 & 1 & \ldots & 1 \\
0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\
0 & 0 & 0 & \ldots & 0 & 1 & 0 & \ldots & 0 \\
0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0
\end{pmatrix}
\]

where \( j \) is output for input \( i \) and 0 otherwise, imply the total-probability conditions for all remaining deterministic operations of the same dimension, i.e.

\[
\forall i_0, j_0, i_1, j_1, \ldots, i_{n-1}, j_{n-1} \geq 0: \text{Tr} (\hat{E} (\hat{D}_{i_0 j_0} \otimes \ldots \otimes \hat{D}_{i_{n-1} j_{n-1}})) = 1 \implies \forall \hat{D}_0, \ldots, \hat{D}_{n-1} \in \Delta: \text{Tr} (\hat{E} (\hat{D}_0 \otimes \ldots \otimes \hat{D}_{n-1})) = 1.
\]

**Proof.** We restrict ourselves to the single-party scenario—the multi-party case follows through linearity. Let \( \hat{v}_i \) be the \( d \)-dimensional vector with a 1-entry at position \( i \) and 0’s everywhere else. We can write a \( d \)-dimensional matrix \( \hat{D}_{ij} \) as

\[
\hat{D}_{ij} = \sum_{m=0}^{d^2} \hat{v}_i \hat{v}_j^T + \hat{v}_i \hat{v}_j^T.
\]

A general deterministic matrix \( D \in \Delta \) of the same dimension, where \( k \) is mapped to \( a_k \), is expressed as

\[
D = \sum_k \hat{v}_k \hat{v}_k^T.
\]

On the one hand, using the antecedent above, we have

\[
\text{Tr} \left( \hat{E} \sum_{k:a_k \neq 0} \hat{D}_{k:a_k} \right) = \sum_{k:a_k \neq 0} \text{Tr} (\hat{E} \hat{D}_{k:a_k}) = \ell
\]

(12)
with \( \ell' = |\{ k | a_k \neq 0 \}|. \) On the other hand, we can rewrite \( \sum_{k:a_k > 0} \hat{D}_{k,a_k} \) as

\[
\sum_{k:a_k > 0} \hat{D}_{k,a_k} = \sum_{k:a_k > 0} \left( \sum_{k_m \neq k} \hat{v}_k \hat{v}_m^T + \hat{v}_a \hat{v}_k^T \right) = \sum_{k:a_k > 0} \hat{v}_k \hat{v}_k^T - \sum_{k:a_k = 0} \hat{v}_k \hat{v}_k^T = D + \sum_{k:a_k > 0} \hat{v}_k \hat{v}_k^T - \sum_{k:a_k = 0} \hat{v}_k \hat{v}_k^T = D + (\ell' - 1) \hat{v}_k \hat{v}_k^T = D + (\ell' - 1) \hat{D}_{0,0}.
\]

Therefore

\[
\text{Tr} \left( \hat{E} \sum_{k:a_k > 0} \hat{D}_{k,a_k} \right) = \text{Tr}(\hat{E}D) + (\ell' - 1) \text{Tr}(\hat{E}\hat{D}_{0,0}) = \text{Tr}(\hat{E}D) + \ell' - 1,
\]

which, with the identity (12), implies

\[
\text{Tr}(\hat{E}D) = 1.
\]

The family \( \{ \hat{D}_{i,j} | i, j \in I \} \) of deterministic operations with the set \( I = \{ 0, \ldots, n-1 \} \) has size \( d(d-1) + 1 \).

**Theorem 6.** (Polytope). The H-representation of the polytope of logically consistent classical processes without predefined causal order is

\[
\forall \hat{D}_0, \hat{D}_1, \ldots, \hat{D}_{n-1} \in \{ \hat{D}_{i,j} | i,j \in I \} : \text{Tr}(\hat{E}(\hat{D}_0 \otimes \hat{D}_1 \otimes \cdots \hat{D}_{n-1})) = 1,
\]

\[
\forall i,j : \hat{E}_{i,j} \geq 0.
\]

The polytope has \( d^{2n} \) facets and dimension

\[
d^{2n} - (d(d-1)+1)^n,
\]

which is exponential in the number of parties.

4. The deterministic-extrema polytope

**Definition 5.** (Deterministic-extrema polytope). The **deterministic-extrema polytope** is defined as the polytope of logically consistent processes without predefined causal order **where all extremal points are deterministic processes** (see polytope with the solid lines in figure 2).

The deterministic-extrema polytope excludes **proper mixtures** of logically inconsistent processes. Such mixtures (consistent mixture of inconsistent points) are convex combinations of deterministic points where at least one deterministic point is outside of the polytope. To find this polytope, one can first solve the extremal points of the general polytope and thereafter select the boolean solutions. These boolean solutions form the \( V \)-representation of the polytope in discussion.

4.1. Three parties, binary inputs, and binary outputs

We discuss the deterministic-extrema polytope in the setting of three parties and binary inputs and outputs. To simplify the presentation, we use \( A = \{ 0 \} \), \( B = \{ 0, 1 \} \), \( C = \{ 0, 1 \} \), \( O_0 = \{ 0, 1 \} \), \( O_1 = \{ 0, 1 \} \), \( O_2 = \{ 0, 1 \} \), \( I_A = \{ 0, 1 \} \), \( I_B = \{ 0, 1 \} \), and \( I_C = \{ 0, 1 \} \). As described in section 3.4, this polytope has 744 extremal points. They can be characterized as follows.

Assume that \( \hat{E} \), when the parties locally apply the identity operation, maps \( (0, 0, 0) \) to \( (0, 0, 0) \), i.e., \( (0, 0, 0) \) is a fixed-point. Then, any other extremal point \( \hat{E}' \) is obtained by the local operations identity \( \hat{D}_3 \) and bit-flip \( \hat{D}_0 \) where we embed these local operations into the environment. Let \( \hat{L}_{a,b,k} \) be the local operation of the three parties.
we look at the case where we assume that exactly two parties exist. Now, we are left with several possibilities on how the parties, i.e., party \( A \) performs the identity if \( i = 0 \) and the bit-flip operation if \( i = 1 \)—the other parties’ local operations are defined in the same way. The extremal point \( \hat{E}' \) can be described as

\[
\hat{E}' = \hat{L}_{ij,k} \hat{\hat{E}} \hat{L}_{ij,k},
\]

where, as described above, the operations are embedded into the environment (see figure 9). Logical consistency of the environment \( \hat{E}' \) follows because we started with a logically consistent \( \hat{E} \) and the operations act on single parties. The process \( \hat{E}' \) maps \((i, j, k)\) to \((j, i, k)\). Thus, starting with \( \hat{E} \), for any choice of \( i, j, k \), we obtain a different extremal point. There are \( 2^3 - 1 \) alternative extremal points that can be constructed in this fashion. From this we conclude that 744/8 = 93 extremal points are such, that \((0, 0, 0)\) is a fixed-point under locally applying the identity. We restrict our analysis to these 93 extremal points; all others can be obtained by the above construction. The following analysis is structured depending on the number of parties that receive a constant from the environment.

There exists only one extremal point where \((0, 0, 0)\) is mapped to \((0, 0, 0)\) under applying identity locally, and where every party receives a constant: the constant \((0, 0, 0)\).

Assume exactly two parties receive a constant, which leaves us with three possibilities of choosing them. Fix these parties to be \( A \) and \( B \). The third party \( C \) receives a value that depends on the operation of \( A \) or of \( B \) or of both. Thus, we are in the case \( A \leq C \) or \( B \leq C \). Inevitably, the constant must be \((0, 0, 0)\); otherwise the fixed-point \((0, 0, 0)\) is not recovered. Party \( C \) receives a value that depends on the value fed back by \( A \) and \( B \); there exist \( 2^3 - 1 = 7 \) such functions where we have excluded the constant and all operations where \( C \) receives a value different from 0 on inputs \((0, 0, 0)\) to the environment from \( A \) and \( B \). Therefore, under all permutations of the parties, 21 extremal points give a constant to two parties and have the fixed-point \((0, 0, 0)\) when the identity is applied locally.

In a next step, assume that exactly one party receives a constant. This assumption, again, allows for three different setups, as we can choose which party receives a constant. Without loss of generality, let \( A \) be this party, i.e., \( A \leq B \) and \( A \leq C \); the constant must be 0 again in order to comply with the requirement of the fixed-point. Now, we are left with several possibilities on how \( B \) and \( C \) depend on \( A \) and on each other. As a first case, we assume that \( B \) and \( C \) do not depend on each other, but depend on \( A \) only \((A \leq B\) and \( A \leq C\)). This dependency cannot be different from the identity channels from \( A \) to \( B \) and from \( A \) to \( C \); the alternative would be the bit-flip channel that would not reproduce the desired fixed-point \((0, 0, 0)\). This gives us three different extremal points under all permutations of the parties. Another possibility on the dependencies is that \( B \) depends on \( A \), and \( C \) depends on \( B \), i.e., \( A \leq B \leq C \), and the interchange of parties \( B \) and \( C \). The channels—by following the same reasoning above—again must be the identity channels; this gives us six extremal points. Now, we look at the case where \( B \) depends on \( A \) and where \( C \) depends on both, \( A \) and \( B \), i.e., \( A \leq B \), \( A \leq C \), and \( B \leq C \), and any permutation of the parties. There are six permutations. The constant that \( A \) receives must be 0, party \( B \) must depend trivially on \( A \) (the identity channel) and party \( C \) can depend in five different ways on \( A \) and \( B \); these are all 2-to-1-bit functions where \((0, 0)\) is mapped to \(0 (2^3) \) minus the constant and minus...
Figure 10. (a) Every party receives a constant. For an environment with fixed point (0, 0, 0) when the parties locally apply the identity map, this constant must be (0, 0, 0). (b) Two parties receive a constant (0, 0), the third party receives a value depending on the other parties’ state fed to the environment. For each of the three cases, seven different functions exist. (c) For each of the three cases, the identity function only is consistent with the setup. (d) Here as well, only the identity channel is consistent with the fixed-point. (e) Five different functions are possible per setup. (f) Here, eight functions per setup are consistent with the fixed-point. (g) No party receives a constant, yet no contradiction arises under any choice of local operations. For a fixed-point (0, 0, 0) where the parties locally apply the identity map, eight different functions that fulfill these requirements exist. In total, the number of deterministic extremal points where (0, 0, 0) is mapped to (0, 0, 0) when the parties apply the identity operation is 1(a) + 3 × 7(b) + 3(c) + 6(d) + 6 × 5(e) + 3 × 8(f) + 8(g) = 93.

the dependencies on $A$ only and on $B$ only. In total, there are $6 \times 5 = 30$ such extremal points. We are left with the last scenario: $B$ depends on $A$ and on $C$, and $C$ depends on $A$ and on $B$, i.e., $A \leq B$, $C \leq B$, $A \leq C$, and $B \leq C$. The constant, as above, is 0. Given the random variable $O_A$, fed to the environment by $A$, the environment can either describe a channel from $B$ to $C$ ($B \leq C$) or describe a channel from $C$ to $B$ ($C \leq B$); any other channel would lead to a causal loop. The direction of the channel must differ under different values $o_A$ fed-back by $C$, as otherwise $B$ and $C$ would not mutually depend on each other. We have two possibilities on the direction given the value fed-back by $A$ is $O_A = 0$. Assume the direction to be $B \leq C$. For the case $O_A = 0$, the channels from $A$ to $B$ and from $B$ to $C$ are the identity channels in order to comply with the fixed point (0, 0, 0). In the other case $O_A = 1$, the direction of the channel between $B$ and $C$ is in the reverse direction compared to $O_A = 0$, i.e., $C \leq B$. Then, because of $O_A \geq 0$, the random variables $I_B$ and $I_C$ are not forced to be (0, 0); there exist two channels from $A$ to $C$ and another two channels from $B$ to $C$. Therefore, we are left with $3 \times 2 \times 4 = 24$ possibilities. An overview over these setups is given in figure 10.

The last setup (see figure 10(g)) where no party receives a constant builds a family of eight extremal points. All extremal points are equivalent up to relabelling of the inputs to and outputs from the environment. One such extremal point is

$$
\hat{E}_{\det} = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 
\end{pmatrix}
$$
The behaviour of this extremal point is

\[ \overline{O} = O \overline{O} O \overline{O} O \overline{O} \]  

where \( \overline{x} \) is the negation of \( x \). This solution is depicted in figure 11. Its location in the polytope is shown in figure 6.

5. Causal games

Given a deterministic extremal point where every party is in the causal past of all other parties, a causal game can be constructed that can be won perfectly in the framework presented here, but that is lost if one assumes a global time.

Definition 6. (Causal game). Let a deterministic process map \( O = (O_0, O_1, \ldots) \) to \( t^O = (t^O_0, t^O_1, \ldots) \), where the \( i \)th entry belongs to party \( S_i \), and where for every \( i \), \( t^O_i \) depends on all other parties’ inputs to the environment. We define a causal game where party \( S_i \) gets a random \( A_i \) and has to produce \( X_i = I^A_i \). The parties are allowed to communicate in a predefined causal order \( [32, 35, 42] \). Let the guesses of all parties be \( X = (X_0, X_1, \ldots) \), and let the random inputs to all parties be \( A = (A_0, A_1, \ldots) \). In a setup with \( n \) parties and where every party obtains and sends a \( d \)-dimensional state, the game’s winning probability is

\[ P_{\text{succ}} = \frac{1}{d^n} \sum_{i} \Pr(X = I^A_i | A). \]  

(14)

Let \( P_{\text{succ}}^C, P_{\text{succ}}^{NC} \) be the success probability of the game (14) with, without the assumption of a predefined causal order, respectively.

Theorem 7. (No winning strategy with predefined order). Using a predefined causal order, the success probability (14) is strictly less than 1, i.e., \( P_{\text{succ}}^C < 1 \).

Proof. For every party \( S_i \), the random variable \( I^A_i \) party \( S_i \) has to guess depends on the other parties’ inputs \( A_{j(i \neq i)} \). In a predefined causal order, however, at least one party is not in the causal future of any other party. Let \( S_i \) be that party, i.e., \( \forall j \ni i: S_i \not\ni S_j \) (see lemma 1). Then, at least for one input \( A_i = a_i \) to \( S_i \), the party \( S_i \) cannot guess perfectly. The success probability \( P_{\text{succ}}^C \) is upper bounded by

\[ P_{\text{succ}}^C = \frac{1}{d^n} \left( \sum_{A_{i,j \neq i}} \Pr(X = I^A_i | A) + \sum_{A_{i,j = a_i}} \Pr(X = I^A_i | A) \right) \]

\[ \leq \frac{1}{d^n} \left( (d-1)d^{n-1} + \sum_{A_{i,j = a_i}} \Pr(X = I^A_i | A) \right). \]

The guessing probability for the non-perfect guess is upper bounded by

\[ \Pr(X = I^A_i | A, A_i = a_i) \leq \frac{d^{n-1} - 1}{d^{n-1}} \]
because for at least one input $A$ with $A_i = a'$, party $S_i$ guesses wrongly. Therefore, we obtain
\[
P_{\text{succ}}^C \leq \frac{1}{d^3} (d - 1)d^{n-1} + d^{n-1}d - 1) \\
\frac{1}{d^3} (d^n - d^{n-1} + d^{n-1} - 1) \\
= 1 - \frac{1}{d^n}.
\]

**Theorem 8.** (Winning strategy without predefined causal order). If we drop the assumption of a predefined causal order, then the causal game (14) can be won perfectly, i.e., $p_{\text{succ}}^{NC} = 1$.

**Proof.** To win the game perfectly, the parties use the process that maps the random variable $O = (O_0, O_1, ...)$ to the random variable $I^O = (I_0^O, I_1^O, ...)$ deterministically, forward their random input to the environment ($A_i = O$), and use the value obtained from the environment as their guess ($X_i = I_i^O$).

For other games, a larger gap between the success probability with a predefined causal order and the success probability without a predefined causal order can be achieved—as is shown in the examples below.

### 6. Examples

We briefly discuss two examples in the three-party scenario. Let $A$, $B$, $C$ be random input bits to the three parties $A$, $B$, $C$, respectively, and let $X$, $Y$, $Z$ be the corresponding output bits.

**Example 3.** An extremal point of the first class of polytopes for three parties and binary inputs/outputs is
\[
\hat{E}_{ext} = \frac{1}{2} \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix};
\]

its behaviour is shown in figure 12. This extremal point is a proper mixture of logically inconsistent processes, as it cannot be written as a convex combination of deterministic points from within the polytope; the left and the right channels from figure 12 individually describe a causal loop and, hence, are logically inconsistent. Initially, this process was used to show that in the classical scenario with three parties or more, correlations without predefined causal order can arise [35]. A causal game that can be formulated for this extremal point is
\[
P_{\text{succ}}^{ex1} = \frac{1}{3} \left( \Pr(X = B \oplus C \mid m = 1) \\
+ \Pr(Y = A \oplus C \mid m = 2) \\
+ \Pr(Z = A \oplus B \mid m = 3) \right),
\]

where, depending on the shared random trit $m$, the party selected by $m$ has to guess the parity of the other two parties’ inputs. If we assume a predefined causal order, then this causal game can be won with probability at most $5/6$ [32]. The reason for this is that at least one party is not in the causal future of the others. This party, hence, can guess the parity with a probability of $1/2$ only. However, by using the environment from figure 12, one can win the game perfectly. To achieve this, if $m = 1$, then party $B$ forwards the random input ($O_B = B$), party $C$ forwards the parity of the random input and the random variable obtained from the environment ($O_C = C \oplus I_C$), and party $A$ uses the random variable obtained from the environment as its guess ($X = I_A$). For the cases $m = 2$ and $m = 3$, the same strategy is used, but where the parties are permuted.

**Example 4.** Another example [36] is depicted in figure 11 and is a deterministic extremal point of the polytope with three parties and binary inputs/outputs (see also equation (13)). Consider the causal game
\[
P_{\text{succ}}^{ex2} = \frac{1}{2} \left( \Pr(X = C, Y = A, Z = B \mid \text{maj}(A, B, C) = 0) \\
+ \Pr(X = \bar{B}, Y = \bar{C}, Z = \bar{A} \mid \text{maj}(A, B, C) = 1) \right),
\]
where \( \text{maj}(A, B, C) \) is the majority of the three bits \( A, B, \) and \( C. \) Whenever the majority of the inputs is 0, i.e., \( \text{maj}(A, B, C) = 0 \), then the parties play the ‘guess-your-neighbours-input’ game [57, 58]: party \( A \) guesses the input of party \( B, \) party \( B \) guesses the input of party \( C, \) and finally party \( C \) guesses the input of party \( A; \) the game is won if all guesses are correct simultaneously. If the majority of the inputs is 1, then they play the same game in reverse direction and flip the output bits. The success probability of winning this game in a world with a predefined causal order is upper bounded by \( \frac{3}{4}. \) This can be seen by the following reasoning. In a predefined causal order, at least a single party has to make a guess without learning anything from the other parties. For example, if party \( A \) causally precedes \( B \) and \( C, \) i.e., \( (A \preceq B \text{ and } A \preceq C), \) then party \( A \) at best always outputs 0 (see table 1). By making such a guess, however, in 2 out of 8 cases, the parties will loose the game, yielding an upper bound of \( \frac{3}{4} \) to the success probability. The same upper bound is achieved by choosing party \( B \) or party \( C \) as causally preceding the others.

By using the environment shown in figure 11, the game can be won perfectly. The parties simply forward their inputs to the environment and use the bits obtained from the environment as the guesses.

### 7. Conclusion and open questions

We describe the polytope for classical multi-party processes without predefined causal order but where the arising correlations are logically consistent. We also describe the polytope formed by deterministic extremal points; it excludes processes that are proper mixtures of logically inconsistent processes, i.e., processes that cannot be written as a convex combination of deterministic processes from within the polytope. For three parties or more, these polytopes contain processes that cannot be simulated by using a predefined causal order among the parties—this is shown by violations of so-called causal inequalities.

A representation with polytopes helps for finding new causal games as well as for optimizing the processes for winning causal games; the optimization problem can be stated as a linear program.

In comparison, it has been shown that the set of causal correlations, i.e., correlations with predefined causal order, also forms a polytope [42, 43]. A complete characterization for the two-party case is given [43], however, in the multi-party case, a characterization is missing. Such a characterization is interesting as then one could subtract it from the polytope studied in this work; this yields an exact characterization of the non-causal processes. Another open question is to decide for which causal games the quantum correlations without causal order outperform their classical counterpart.

### Acknowledgments

We thank Mateus Araújo, Veronika Baumann, Cyril Branciard, Časlav Brukner, Fabio Costa, Adrien Feix, Arne Hansen, Alberto Montina, and Benno Salwey for helpful discussions. Furthermore we thank the anonymous
referees for the detailed comments. The present work was supported by the Swiss National Science Foundation (SNF) and the National Centre of Competence in Research ‘Quantum Science and Technology’ (QSIT).

References

[1] Hardy L 2007 J. Phys. A: Math. Theor. 40 3081
[2] Einstein A, Podolsky B and Rosen N 1935 Phys. Rev. 47 777
[3] Bell J S 1964 Physics 1 195
[4] Brunner N, Cavalcanti D, Pironio S, Scarani V and Wehner S 2014 Rev. Mod. Phys. 86 419
[5] Reichenbach H 1956 (The Direction of Time) ed M Reichenbach (Berkeley: University of California Press) ch 19, pp 157–167
[6] Suarez A and Scarani V 1997 Phys. Lett. A 232 9
[7] Stefanov A, Zbinden H, Gisin N and Suarez A 2002 Phys. Rev. Lett. 88 120404
[8] Scarani V and Gisin N 2002 Phys. Lett. A 295 167
[9] Coretti S, Haagerup I and Wolf S 2011 Phys. Rev. Lett. 107 100402
[10] Bancal J-D, Pironio S, Acin A, Liang Y-C, Scarani V and Gisin N 2012 Nat. Phys. 8 867
[11] Barnea T, Bancal J-D, Liang Y-C and Gisin N 2013 Phys. Rev. A 88 022123
[12] Scarani V, Bancal J-D, Suarez A and Gisin N 2014 Found. Phys. 44 523
[13] Wood C J and Spekkens R W 2015 New J. Phys. 17 033002
[14] Leibniz G W and Clarke S 2000 (Correspondence) ed R Ariew (Indianapolis: Hackett Publishing Company, Inc.) pp 14–18
[15] von Weizsäcker C F 1985 Aufbau der Physik (München: Carl Hanser Verlag) p 522
[16] Page D N and Wootters W K 1983 Phys. Rev. D 27 2885
[17] Wootters W K 1984 Int. J. Theor. Phys. 23 701
[18] Bombelli L, Lee J, Meyer D and Sorkin R D 1987 Phys. Rev. Lett. 59 521
[19] D’Ariano G M 2011 AIP Conf. Proc. 1363 63
[20] Erker P 2014 The quantum hourglass Master’s Thesis Zürich ETH
[21] Vedral V 2014 in preparation (arXiv:1408.6965 [quant-ph])
[22] Giovannetti V, Lloyd S and Maccone L 2015 Phys. Rev. D 92 045033
[23] Rankovic S, Liang Y-C and Renner R 2015 in preparation (arXiv:1506.01373 [quant-ph])
[24] Colbeck R and Renner R 2011 Nat. Commun. 2 411
[25] Ghirardi G and Romano R 2013 Found. Phys. 43 881
[26] Hardy L 2005 in preparation (arXiv:0509120 [gr-qc])
[27] Hardy L 2009 Quantum gravity computers: on the theory of computation with indefinite causal structure Quantum Reality, Relativistic Causality, and Closing the Epistemic Circle (The Western Ontario Series in Philosophy of Science vol 73) (Netherlands: Springer) pp 379–401
[28] Chiribella G, D'Ariano G M and Perinotti P 2009 Phys. Rev. A 80 022339
[29] Chiribella G 2012 Phys. Rev. A 86 040301
[30] Colnaighi T, D’Ariano G M, Facchinì S and Peri-notti P 2012 Phys. Lett. A 376 2940
[31] Oreshkov O, Costa F and Brukner Č 2012 Nat. Commun. 3 1092
[32] Baumeler A and Wolf S 2014 Proc. Int. Symp. on Information Theory (Piscataway, NJ: IEEE) pp 526–530
[33] Chiribella G, D’Ariano G M, Perinotti P and Valiron B 2013 Phys. Rev. A 88 022318
[34] Costa F 2013 Local and causal structures in quantum theory PhD Thesis Universität Wien, Wien
[35] Baumeler A, Feix A and Wolf S 2014 Phys. Rev. A 90 042106
[36] Brukner Č 2015 New J. Phys. 17 083034
[37] Ihnousséin I and Grinbaum A 2015 Phys. Rev. A 92 042124
[38] Ihnousséin I 2014 Quantum correlations and causal structures PhD Thesis Université Paris-Sud
[39] Morimae T 2014 Phys. Rev. A 90 010101
[40] Oreshkov O and Cerf N J 2014 in preparation (arXiv:1406.3829 [quant-ph])
[41] Araújo M, Branciard C, Costa F, Feix A, Giarmatzi C and Brukner Č 2015 New J. Phys. 17 102001
[42] Oreshkov O and Giarmatzi C 2015 in preparation (arXiv:1506.05449 [quant-ph])
[43] Branciard C, Araújo M, Feix A, Costa F and Brukner Č 2016 New J. Phys. 18 013008
[44] Feix A, Araújo M and Brukner Č 2015 Phys. Rev. A 92 052326
[45] Oreshkov O and Cerf N J 2015 Nat. Phys. 11 853
[46] Brukner Č 2014a Nat. Phys. 10 259
[47] Brukner Č 2014b private communication
[48] Price H 1991 Br. J. Phil. Sci. 42 157
[49] Woodward J 2003 Making Things Happen: A Theory of Causal Explanation (Oxford: Oxford University Press)
[50] Barjavel R 1944 Le Voyageur Imprudent (Paris: Denoël)
[51] Jamiołkowski A 1972 Rep. Math. Phys. 3 275
[52] Choi M-D 1975 Linear Algebra Appl. 10 285
[53] Motzkin T S, Raiffa H, Thompson G L and Thrall R M 1953 Contributions to Theory of Games (Annals of Mathematics Studies vol 2) ed H Wuhn and A W Tucker (Princeton, NJ: Princeton University Press)
[54] Fukuda K and Prodon A 2003 Combinatorics and Computer Science ed M M Deza et al (Berlin: Springer) pp 91–111
[55] Fukuda K 2003 ‘cdd_home’ (http://www.ifor.isy.slu.se/personal/fukudak/cdd_home/cdd.html)
[56] Araújo M and Feix A 2014 private communication
[57] Almeida M L, Bancal J-D, Brunner N, Acín A, Gisin N and Pironio S 2010 Phys. Rev. Lett. 104 230404
[58] Acín A, Almeida M L, Augusiak R and Brunner N 2016 Guess your neighbour’s input: no quantum advantage but an advantage for quantum theory Quantum Theory: Informational Foundations and Foils (Netherlands: Springer)