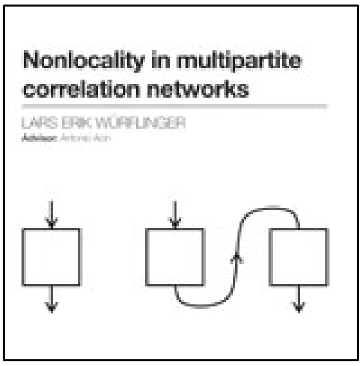
Nonlocality in multipartite correlation networks

LARS ERIK WÜRFLINGER

Advisor: Antonio Acín



ICFO-The Institute of Photonic Sciences

1. Introduction

Since its formulation in the 1920's quantum mechanics has become the probably most successful and thoroughly tested physical theory. The mathematical formalism of quantum mechanics, developed as a response to the failure of the then-existing classical theories to explain phenomena such as black body radiation or the discrete spectra of atoms, is today routinely applied to predict results of measurements with remarkable accuracy in many branches of modern physics.

Despite the success of the theory in predicting the outcomes of experiments and the consensus among physicists concerning how the quantum-mechanical rules should be applied, the conceptual foundations of quantum mechanics have been a subject of research and debate since the early days of the theory. The non-classical phenomena of entanglement and nonlocality were what led Einstein et al. (1935) to express their unease with the theory and consider the quantum-mechanical description as "incomplete".

Formally, entangled states are a direct consequence of the way quantum mechanics describes composite systems. At the same time they are at the heart of the struggle with quantum mechanics, as their behaviour presents a dramatic departure from classical physics: even if the spatial components of a composite physical system are separated and brought to locations arbitrarily far from each other, the response of one component when subjected to a measurement may still be affected by actions performed on the other component. This sounds as if the two parts could communicate instantaneously, but the rules of quantum mechanics guarantee that this nonlocal "action at a distance" cannot be used for communication. Correlations like these that do not allow for communication are said to fulfil the no-signalling principle.

In 1964 Bell reassessed the argument presented in (Einstein et al., 1935). He was able to formulate the ideas of classicality and locality in clear mathematical assumptions, which allowed him to prove that no local classical theory can explain the behaviour predicted by quantum mechanics (Bell, 1964). It is hard to underrate the importance of this result, as it allows one to falsify the way physical theories were built for ages in classical physics.

With more efficient sources for entangled states becoming available, falsifiable criteria for local classical theories, derived from Bell's assumptions and expressed as inequalities (Clauser et al., 1969), could be put to experimental tests. The first reliable violation of the inequalities, as predicted by quantum mechanics, was achieved in an experiment by Aspect et al. (1982). Since then numerous Bell tests have been performed confirming the predictions of quantum mechanics, hence building the case for entanglement and nonlocality.

Not only were entanglement and nonlocality verified experimentally, but in the last decades both these properties were also identified as useful resources for information processing, giving birth to the field of quantum information theory. This field studies the implications of quantum mechanics on the way information can be stored and processed. Using quantum systems to encode and manipulate information new information processing protocols become possible, such as efficient integer factorisation (Shor, 1994) or secure quantum cryptography (Bennett and Brassard, 1984; Ekert, 1991).

The importance of entanglement as a resource for quantum information has driven a strong theoretical effort devoted to its characterisation, detection and quantification (Horodecki et al., 2009). Many new mathematical tools that resulted from the study of entanglement, such as entanglement witnesses or entanglement measures, also find application beyond the field of quantum information for which they were initially developed, e.g. in condensed matter physics (Osterloh et al., 2002) or quantum thermodynamics (Popescu et al., 2006).

Recently, a new paradigm was introduced in the field of quantum information: device-independent quantum information processing (Barrett and Pironio, 2005; Acín et al., 2007; Pironio et al., 2010; Colbeck and Kent, 2011; Masanes et al., 2011). There, the main goal is to achieve an information processing task without making any assumptions about the internal working of the devices used in the protocol. This device-independence makes such applications appealing, both from a theoretical and practical viewpoint.

In this scenario, the objects of interest are correlated systems distributed among several observers. Each observer can choose a classical variable as input for his system, which produces a classical output. The system is just seen as a black box and no assumption is made about the internal process producing the output given the input, except that it cannot contradict quantum theory. The observed correlations among the input-output processes of each system are described by joint conditional probability distributions. The existence of nonlocal quantum correlations opens the possibility for information processing tasks with no classical counterpart.

The approach of device-independence in quantum information leads to the identification of nonlocality as an information resource, alternative to entanglement. Even though the only known way of generating nonlocal correlations among different observers consists in measuring entangled quantum states, it is a well-established fact that entanglement and nonlocality represent inequivalent properties (Acín et al., 2002; Methot and Scarani, 2007). Thus, given the success of entanglement theory, it is desirable to have an analogous theory for the resource of nonlocality.

This thesis sets out to develop such a theoretical framework for the characterisation of nonlocality as a resource. As we will see, it is necessary to study situations more general than the scenario originally considered by Bell (1964) to gain a better understanding of the phenomenon of nonlocality. By investigating scenarios of several parties distributed in network-like structures, this thesis provides new descriptions of the resource of nonlocality. These findings also have implications for the general characterisation of quantum correlations and the detection of new forms of nonlocality.

Before we can address all these questions, we present the general notion of a correlation scenario in the introductory Chapter 2 along with other fundamental concepts and definitions that will be used in the remainder of this thesis. Our main results are contained in Chapters 3 to 6 and can be summarised as follows.

In Chapter 3 we tackle the question of how nonlocality can be defined consistently in a scenario of arbitrarily many parties where collaboration among some of them is allowed. To this end we need to identify the allowed operations for this physical situation; then nonlocality is defined as the resource that cannot be created by these operations. As it turns out, the standard definition of multipartite nonlocality, adopted by the community so far, is inconsistent with our operational characterisation. Therefore, we introduce a new class of models that overcome these inconsistencies.

By using a special class of these models we show in Chapter 4 that our findings have implications for the characterisation of the set of quantum correlations. Information principles were recently proposed as a means to single out this set from the larger set of correlations that are only constrained by the no-signalling principle. We can show that any such principle that aims to achieve this task must be genuinely multipartite.

We then developed a description of nonlocality in an even more generalised scenario of several parties in Chapter 5. There, the parties are allowed to perform not single but sequences of measurements on their systems. Characterising nonlocality also in this scenario in operational terms and defining local models compatible with this definition, we show that a new form of nonlocality can arise.

Lastly, in Chapter 6, we examine the problem of detecting the presence of nonlocality in a multipartite scenario when one is given only partial access to the global system. We find that one can verify that the total system must display nonlocality, even though the accessible subsystems only exhibit local correlations.

The final Chapter 7 concludes this thesis by summarising the results and placing them in a broader context. Further, some open question concerning the presented work and future research perspectives are outlined.

2. Correlation scenarios

This chapter sets the stage. We introduce the general idea of a correlation scenario that will be studied from different perspectives in the following chapters. We fix some basic notation and provide the most important concepts, such as quantum and nonlocal correlations, and Bell inequalities.

2.1. General notation

A measurement is the assignment of an output to a physical object by using an instrument or device. In a situation, where several measurements are performed at different sites on the same physical system, we are interested in the correlations between the outputs obtained by the different measurements. It is useful to think of the different measurements being performed by different parties, where each party receives its part of the total system that is produced by a common source.

Let us first consider the bipartite case. The scenario is characterised by specifying the sets of possible measurement devices, which we will also refer to as inputs, and the corresponding outputs for the two parties A and B. In every run of the experiment the source produces a physical system, each party receives a part of that system, chooses a measurement to perform and records the obtained output. After many runs of the experiment the parties can come together and assign probabilities to the different events using relative frequencies, see Fig. 2.1. Mathematically, the object of interest is the joint probability distribution for the outcomes given the measurement devices. We will write

$$P(ab|xy) \tag{2.1}$$

to denote such an observed joint probability of party A obtaining result a when using device x and party B obtaining result b when using device y. The collection of all these joint conditional probabilities will be called the *correlations* of the given scenario, where we assume for simplicity that the sets of possible outcomes and measurements are finite for both A and B.

As we want to interpret the numbers $P(ab|xy)$ as probabilities, we have the

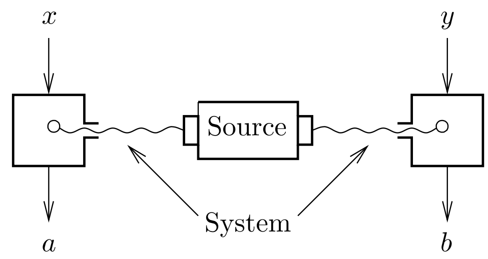


Figure 2.1.: Bipartite correlation scenario. In every run of the experiment a common source prepares a physical system and each of the two parties receives a subsystem. The parties A and B choose their measurement settings x and y respectively and observe the outcomes a and b. After many runs of the experiment the parties get together and calculate the correlations, i.e. the joint probabilities $P(ab|xy)$ of observing the outcomes a for A and b for B given the measurement settings x and y.

obvious conditions $P(ab|xy) \geq 0$ and

$$\sum_{a,b} P(ab|xy) = 1 \tag{2.2}$$

for all measurements x, y. Sometimes we will also be interested in the *marginal distributions* of the parties. They correspond to the probabilities observed by one party alone, i.e.

$$P_A(a|xy) = \sum_b P(ab|xy) \tag{2.3}$$

$$P_B(b|xy) = \sum_a P(ab|xy). \tag{2.4}$$

In general, a marginal distribution, say for A, may depend not only on the measurement x chosen by A but also on the measurement choice y by B. If this was the case, B could use this dependence to communicate a message to A just by the local choice of his measurement setting y. Sending a message, however, will always require some physical system travelling from B to A, which does not correspond to the scenario we want to consider: the common source distributes the subsystems to the parties and no communication takes place between A and B. Thus, we further require that the correlations fulfil the *no-signalling principle*, i.e.

$$\begin{aligned} &\sum_b P(ab|xy) \quad \text{independent of } y \\ &\sum_a P(ab|xy) \quad \text{independent of } x. \end{aligned} \tag{2.5}$$

In other words, no-signalling means that the objects

$$P(a|x) = \sum_b P(ab|xy) \tag{2.6}$$

$$P(b|y) = \sum_a P(ab|xy) \tag{2.7}$$

are well defined. Note, however, that the above notation introduces some ambiguity, as it is not clear whether an expression like $P(o|m)$ refers to the marginal distribution of A or B when m is a valid label for a measurement setting and o a possible outcome for both A and B. To avoid this ambiguity one could introduce additional subscripts as in $P_A(a|x)$ to indicate that the expression refers to the marginal distribution of A, a notation we will use in a few cases. In most cases, however, the ambiguity will be resolved by either context or the use of suggestive labels a, x and b, y as above. Another notational issue concerns the sets of measurement choices and their corresponding outcomes. Throughout this work, with a few exceptions, we will not make these sets explicit but only implicitly assume that they are finite. The results obtained in these cases are valid for all correlations scenarios with finite sets of measurements and outputs.

As mentioned in the introduction, this thesis studies nonlocal correlations in scenarios that go beyond the standard bipartite case originally studied by Bell. Let us therefore generalise the considerations we made so far to the case of more than two parties. Thus, for the case of n parties labelled $A_1, \dots, A_n$ the correlations are the collection of the joint probabilities

$$P(a_1 \dots a_n | x_1 \dots x_n) \tag{2.8}$$

for the outputs $a_1, \dots, a_n$ given the inputs $x_1, \dots, x_n$. We assume the same physical situation as in the bipartite case in which a common source distributes the subsystems to the parties and no communication takes place between any of the parties. Then the the no-signalling condition states that for all $1 \leq i \leq n$

$$\sum_{a_i} P(a_1 \dots a_n | x_1 \dots x_n) \quad \text{is independent of } x_i, \tag{2.9}$$

which means that the marginal probabilities for every group of parties are well defined. One can think of such a collection of probabilities as one large device with n slots for the inputs and n pointers indicating the output of a given measurement. Due to the no-signalling condition, every party A_i observes the outcome a_i for the given measurement x_i with the probability

$$P_{A_i}(a_i|x_i) = \sum_{\{a_j|a_j \neq a_i\}} P(a_1 \dots a_n|x_1 \dots x_n), \tag{2.10}$$

independent of the measurements performed by the other parties. Interpreting the correlations as this big input-output device shared by n parties, we will also refer to them as an *n-partite nonsignalling box P*.

2.2. Quantum correlations

The previous section, that defines general correlation scenarios, does not refer to a specific way physical systems and measurements on them are described. This section introduces a special kind of correlations, namely those that arise in a correlation scenario if one describes physical systems and measurements according to the formalism of quantum mechanics.

A quantum system is specified by a Hilbert space $\mathfrak{H}$ and a linear map $\varrho : \mathfrak{H} \to \mathfrak{H}$, called the state of the system; we only consider the case of finite-dimensional Hilbert spaces, in which case $\mathfrak{H} = \mathbb{C}^d$. The state is a semi-definite positive matrix and has unit trace, i.e. $\varrho \geq 0$ and $\operatorname{tr} \varrho = 1$.

A general measurement x on the system is given by a *positive-operator valued measure* (POVM), i.e. an assignment $a \mapsto M_{a|x}$ for every outcome a of the measurement to a semi-definite positive operator $M_{a|x} \geq 0$ on $\mathfrak{H}$ such that

$$\sum_a M_{a|x} = \mathbb{I}_d, \tag{2.11}$$

where $\mathbb{I}_d$ denotes the d-dimensional identity matrix.

The probability $P(a|x)$ of obtaining the outcome a when using the measurement device x is then given by the the Born rule

$$P(a|x) = \operatorname{tr}(\varrho M_{a|x}). \tag{2.12}$$

The conditions on $M_{a|x}$ together with the trace rule guarantee positivity and normalisation of $P(a|x)$. This general notion of measurements includes the special case of a *projector-valued measure*, where for each measurement x the outputs are assigned to orthogonal projectors acting on $\mathfrak{H}$, i.e. $a \mapsto E_{a|x}$, where $E_{a|x}$ self-adjoint, $E_{a|x}E_{a'|x} = \delta_{aa'}E_{a|x}$ and $\sum_a E_{a|x} = \mathbb{I}_d$. In the case of a projector-valued measure one can define the *post-measurement state*, the state the system is left in after the outcome a has been obtained when performing the measurement x, as

$$\varrho_{a|x} = \frac{E_{a|x}\varrho E_{a|x}}{\operatorname{tr}(\varrho E_{a|x})}. \tag{2.13}$$

When considering a system composed of n individual systems with Hilbert spaces $\mathfrak{H}_1, \dots, \mathfrak{H}_n$, the total system is described by the tensor product space $\mathfrak{H} = \bigotimes_i \mathfrak{H}_i$ and a state $\varrho : \mathfrak{H} \to \mathfrak{H}$. For every party i the measurement x_i is given by a POVM $\{M^{(i)}_{a_i|x_i}|a_i\}$ on $\mathfrak{H}_i$ and the joint probabilities are calculated according to

$$P(a_1 \dots a_n|x_1 \dots x_n) = \operatorname{tr}(\varrho M^{(1)}_{a_1|x_1} \otimes \dots \otimes M^{(n)}_{a_n|x_n}). \tag{2.14}$$

The conditions on the POVM elements imply that such correlations fulfil the no-signalling principle and correlations that can be written in the above form are called *quantum correlations*. Given an n-partite nonsignalling box it is in general hard to decide whether it has a quantum representation as in Eq. (2.14). For the case of two dichotomic projective measurements the problem was solved by Fritz (2010); the possible undecidability of the general problem was discussed in (Wolf et al., 2011).

Navascués et al. (2007, 2008) introduced an infinite hierarchy of conditions that must be satisfied by a nonsignalling box to have a quantum representation. Given nonsignalling correlations that do not have a quantum realisation, such correlations will be certified as non-quantum at some finite level of the hierarchy. Another approach to characterise the set of quantum correlations is to use concepts from information theory, a subject we will get back to in Chapter 4.

Another aspect of quantum correlations is the type of correlations contained in the mathematical structure of the quantum state itself. Formally, one calls a state ϱ acting on the composite Hilbert space $\mathfrak{H} = \bigotimes_i \mathfrak{H}_i$ *entangled*, if it is not a convex sum of product states, i.e. if it cannot be written as

$$\varrho = \sum_i p_i \varrho_1^{(i)} \otimes \dots \otimes \varrho_n^{(i)}, \tag{2.15}$$

where the positive p_i sum to unity and $\varrho_j^{(i)}$ is a quantum state on $\mathfrak{H}_j$ for all i. States with a decomposition as in Eq. (2.15) are called separable. More generally, for a partition $\Pi = \{C_1, \dots, C_k\}$ of $\{1, \dots, n\}$ one calls ϱ separable with respect to Π, if it can be written as

$$\varrho = \sum_i p_i \varrho_{C_1}^{(i)} \otimes \cdots \otimes \varrho_{C_k}^{(i)}, \tag{2.16}$$

where $\varrho_{C_j}^{(i)}$ is a quantum state on $\bigotimes_{r\in C_j} \mathfrak{H}_r$ for all i. Finally, a state is called k-separable if it can be written as the convex sum of states, each of which is separable with respect to a partition of $\{1,\ldots,n\}$ into k groups.

Entanglement is obviously something characteristic of quantum mechanics as it is defined in terms of the mathematical structure of the theory unpresent in classical physics. But this quantumness can also manifest itself in a general correlation scenario as defined in the previous section: as Bell (1964) showed, local measurements on entangled quantum states can give rise to correlations that cannot be explained by any local classical theory. The next section will discuss this remarkable fact and make precise what is meant by a local classical theory.

2.3. Bell's theorem and nonlocal correlations

To say that quantum mechanics is not a classical theory is one thing, to say that the correlations displayed by quantum systems in a correlation scenario cannot be explained by any local classical theory is another. This far reaching conclusion was reached in a theorem by Bell (1964).

The unease with quantum mechanics, especially concerning the existence of entangled states, had already been expressed as early as 1935 in the now famous papers by Einstein et al. (1935) and Schrödinger (1935). But it was Bell (1964), who presented the dilemma with quantum theory in the form of clear assumptions. These assumptions, that concern notions of classicality and locality, allowed Bell to exclude a whole class of models as possible explanations for quantum correlations.

To arrive at a formal definition of a local classical model let us go back to the situation of a general correlation scenario, where, for simplicity, we consider for now the bipartite case. There are two parties, A and B, each of them in the possession of some measurement devices. The common source produces physical systems and sends one part of it to A and the other part to B. To characterise the behaviour of the source let us introduce a *hidden variable* λ that takes values in some space Λ. The source is then characterised by a probability measure μ on Λ, i.e. $\int_U \mu(\mathrm{d}\lambda)$ is the probability that a physical system described by λ with $\lambda \in U$ is produced by the source for some (measurable) set $U \subseteq \Lambda$. This variable is to be thought of as a description of the physical system but is itself not observable. One can think of λ as a label attached to the system emitted by the source or as something like the state of the emitted system.

The assumption of locality in Bell's theorem has a clear operational interpretation: the variable λ describes the systems sent to A and B in such a way that it is possible for each party to compute the output, or at least its probability, for every possible measurement choice. Therefore, we get a probability function $\lambda \mapsto P_A^\lambda(a|x)$ for every possible output a and every measurement choice x of A and a similar assignment for B. The correlations of the entire experiment can then be computed by averaging over the variable λ with respect to the probability measure, i.e.

$$P(ab|xy) = \int_\Lambda \mu(\mathrm{d}\lambda)\ P_A^\lambda(a|x) P_B^\lambda(b|y). \tag{2.17}$$

Correlations that can be decomposed in this form are said to admit a *local hidden-variable model* (LHVM). Note that such models automatically fulfil the no-signalling condition. Further, the assumption of locality can also be expressed as the separability condition for the joint probabilities for a given λ:

$$P^\lambda(ab|xy) = P_A^\lambda(a|x) P_B^\lambda(b|y). \tag{2.18}$$

So, in a LHVM one assumes that the response of one party, say A, for a given λ only depends on the choice of measurement x of that party and not on the measurement device used by B.

This clear mathematical formulation of LHVMs allows one to falsify not only a specific model trying to explain certain correlations, but a whole class of theories, namely the very way theories were formulated for centuries in classical physics.

The standard example of a correlation scenario demonstrating that there are quantum correlations that cannot be described by a local hidden-variable model is the Clauser-Horne-Shimony-Holt (CHSH) scenario (Clauser et al., 1969). It considers a scenario where each of two parties has two measurements with two possible outcomes, where we will use $a, b \in \{-1, 1\}$ and $x, y \in \{0, 1\}$ to label the outcomes and measurements. Consider the expectation values for a given λ

$$\langle a_x b_y(\lambda)\rangle = \sum_{a,b} ab\ P^\lambda(ab|xy), \tag{2.19}$$

then the expression

$$\beta(\lambda) = \langle a_0 b_0(\lambda)\rangle + \langle a_0 b_1(\lambda)\rangle + \langle a_1 b_0(\lambda)\rangle - \langle a_1 b_1(\lambda)\rangle \tag{2.20}$$

fulfils $|\beta(\lambda)| \leq 2$ for all λ. Therefore, we also have $|\beta| \leq 2$, where

$$\beta = \int_\Lambda \mu(\mathrm{d}\lambda)\beta(\lambda) \tag{2.21}$$

the expectation of $\beta(\lambda)$. Writing β in terms of the correlators

$$C(x, y) = \sum_{a,b} ab\, P(ab|xy) \tag{2.22}$$

$|\beta| \leq 2$ becomes the famous CHSH inequality

$$|C(0,0) + C(0,1) + C(1,0) - C(1,1)| \leq 2. \tag{2.23}$$

It is possible to violate this inequality with correlations obtained from local measurements on a quantum state. Consider the two-qubit state

$$|\Phi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \tag{2.24}$$

and the measurements $\{M_{a|x}\}$ for A and $\{N_{b|y}\}$ for B, where

$$\begin{aligned} M_{a|0} &= \frac{\mathbb{I}_2 + a\sigma_z}{2}, \quad M_{a|1} = \frac{\mathbb{I}_2 + a\sigma_x}{2}, \\ N_{b|0} &= \frac{\mathbb{I}_2 + b\sigma_+}{2}, \quad N_{b|1} = \frac{\mathbb{I}_2 + b\sigma_-}{2}, \end{aligned} \tag{2.25}$$

and $\sigma_\pm = \frac{1}{\sqrt{2}}(\sigma_z \pm \sigma_x)$. Then, calculating $P(ab|xy) = \langle\Phi| M_{a|x} \otimes N_{b|y} |\Phi\rangle$ one finds for the CHSH expression

$$C(0,0) + C(0,1) + C(1,0) - C(1,1) = 2\sqrt{2}. \tag{2.26}$$

This shows that these quantum correlations cannot be explained by a local hidden-variable model for the given scenario. The expression Eq. (2.23) is the most prominent example of what is called a *Bell inequality*, an inequality fulfilled by all correlations admitting a local hidden-variable model in a given correlation scenario. Correlations that fulfil all Bell inequalities of a given scenario are called local, whereas any correlations violating at least one Bell inequality are called *nonlocal*.

The CHSH scenario is certainly the best studied correlation scenario and Tsirelson (1983) showed that for all quantum states and measurements, i.e. without any restriction on the dimensions of the local Hilbert spaces, the optimal value for the CHSH expression is given by $2\sqrt{2}$. However, if one does not restrict the study to quantum correlations but allows for arbitrary nonsignalling correlations, higher values for the CHSH expression can be obtained. Popescu and Rohrlich (1994) showed that the following bipartite nonsignalling box with binary inputs, $x, y \in \{0,1\}$ and outputs $a, b \in \{0,1\}$

$$P_{\mathrm{PR}}(ab|xy) = \begin{cases} \frac{1}{2} & a+b \equiv xy \pmod 2 \\ 0 & \text{otherwise} \end{cases} \tag{2.27}$$

leads to a CHSH value of 4, the algebraic maximum of the expression. This box, also called PR-box, is an example of an extremal nonsignalling distribution. The bound of $2\sqrt{2}$ for quantum systems and the existence of the PR-box show that the set of classical correlations is strictly contained in the set of quantum correlations, which, in turn, is strictly contained in the set of nonsignalling correlations.

Now, it seems natural to ask how the concept of a local hidden-variable model can be generalised to the case of n parties. In the discussion on quantum states in Section 2.2 we have seen that the notion of separability of an n-partite system is in general defined with respect to some partition Π of $\{1, \ldots, n\}$. The question how one can define locality of n-partite correlations with respect to a partition in a consistent manner will precisely be the subject of Chapter 3.

To conclude this section, let us just say that the generalisation of local hidden-variable models to more than two parties is straightforward when considering the partition $\Pi = \{1|2|...|n\}$. For this case an n-partite nonsignalling box P is said to admit a LHVM, if for a space Λ with probability measure μ there is an assignment $\lambda \mapsto P_i^\lambda(a_i|x_i)$ to probability functions of the outcomes a_i given the measurement x_i for all $1 \leq i \leq n$, such that

$$P(a_1 \ldots a_n|x_1 \ldots x_n) = \int_\Lambda \mu(\mathrm{d}\lambda)\; P_1^\lambda(a_1|x_1) \ldots P_n^\lambda(a_n|x_n). \tag{2.28}$$

2.4. Bell inequalities and convex geometry

In Bell's formulation of local hidden-variable models, as in Eq. (2.28), the response functions P_i^λ in general only allow one to compute the probability for the measurement outcome. A further requirement for a local classical theory would be *determinism*, i.e. to demand that the response function only take values in $\{0,1\}$. As in turns out, however, this requirement does not lead to stronger restrictions than those imposed by the original probabilistic LHVM: if the correlations P admit a local hidden-variable model as in Eq. (2.28), then P also allows for a *deterministic local hidden-variable model*, where the response functions P_i^λ take values in $\{0,1\}$ for $1 \leq i \leq n$ and all λ (Fine, 1982).

In a deterministic LHVM, for λ and x_i given, the function $P_i^\lambda(a_i|x_i)$ will attain unity for one specific outcome and vanish for all other outcomes. If one now considers an n-partite correlation scenario, where every party can choose from m measurements that each can give r different outcomes, then there is a total of nm measurements. Thus, for a given λ in a deterministic model one can now assign to each of the nm measurements exactly one of the possible r outcomes. We call such an assignment a deterministic strategy. In other words, the space Λ of the hidden variable is made up of r^{nm} pieces, where each piece is characterised by a deterministic strategy. This permits to write Eq. (2.28) as a sum

$$P(a_1 \dots a_n|x_1 \dots x_n) = \sum_s p_s P_1^s(a_1|x_1) \dots P_n^s(a_n|x_n), \qquad (2.29)$$

where the summation is over the r^{nm} deterministic strategies assigning one of the r outputs to every of the nm measurements and p_s is the probability of the strategy s, i.e. $p_s = \int_{U_s} \mu(\mathrm{d}\lambda)$ for the region U_s of Λ corresponding to the strategy s.

So far we have only seen one example of a Bell inequality, namely the CHSH inequality, which corresponds to the case $(n, m, r) = (2, 2, 2)$. If we consider the general case of n parties with m measurements and r outcomes for each measurement, we are dealing with m^n different measurement settings and r^n different outcomes. Thus, we have a total of $(rm)^n$ probabilities and are confronted with the problem to find inequalities that demarcate the set of correlations that can be obtained within a local hidden-variable model from those that are not compatible with such a model. Ignoring the constraints given by normalisation and no-signalling, one can think of the probabilities as vectors v from a $(rm)^n$-dimensional space. Now, the above analysis of deterministic models tell us that every such vector can be written as the convex sum of at most $k = r^{nm}$ probability vectors v_s given by the deterministic strategies. Hence, the set $\mathfrak{L}$ of local correlations is the convex hull

$$\mathfrak{L} = \operatorname{conv}\{v_1, \dots, v_k\} = \{p_1 v_1 + \ldots + p_k v_k | p_i \geq 0, \sum_i p_i = 1\} \qquad (2.30)$$

of the extremal points $\{v_s\}$. Since the number of extremal points is finite, $\mathcal{L}$ is a *convex polytope*.

Every convex polytope can be either described by the convex hull of its extremal points, the *V-description*, or as the intersection of a finite number of half-spaces, the *H-description*. In general, the definition as the intersection of half-spaces does not imply that the corresponding set is bounded. The set $\mathfrak{L}$ of local correlations, however, is bounded and the inequalities that describe the half-spaces of its H-description are the Bell inequalities for the given correlation scenario. Thus, the problem of finding the Bell inequalities for a given scenario is equivalent to finding the H-description of the convex polytope $\mathfrak{L}$ given its V-description, i.e. given Eq. (2.30) find vectors $\beta_1, \ldots, \beta_l$ such that

$$\mathfrak{L} = \{v | \beta_i \cdot v \leq 1, i = 1, \ldots, l\}. \tag{2.31}$$

So we know that there is a finite number of Bell inequalities for every correlation scenario, but finding a complete description of the local polytope in terms of inequalities for a general scenario is computationally hard (Pitowsky, 1989; Werner and Wolf, 2001) and a general solution is unlikely to exist. Therefore, in practice one either restricts the investigation to small values of (n, m, r) or cases with additional symmetries.

3. Operational framework for nonlocality

Both entanglement and nonlocal correlations are not only characteristic features of quantum theory, but they also constitute important resources for information processing. Identification of entanglement as a resource for quantum information processing has led to an alternative characterisation of entanglement: instead of defining it merely formally, as done in Section 2.2, entanglement can also be defined as a property of composite quantum states that cannot be created by a certain class of operations, which captures the role of entanglement as a resource (see e.g. the review by Horodecki et al., 2009).

Within the recently introduced framework of device-independent quantum information processing also nonlocality has been identified as a new quantum resource for information processing (Barrett and Pironio, 2005; Acín et al., 2007; Pironio et al., 2010; Colbeck and Kent, 2011; Masanes et al., 2011). There, the main goal is to achieve an information processing task without making any assumptions about the internal working of the devices used in the protocol. The device-independence of these applications makes them appealing, from the viewpoint of both theory and implementation.

Motivated by the success of the operational approach to characterise entanglement and given the fact that nonlocality is known to be inequivalent to entanglement, we set out to develop an analogous operational framework for the resource of nonlocality.

3.1. Operational definition of entanglement

This section reviews the operational definition of entanglement to illustrate the idea how one formulates a resource theory. The first step when deriving such an operational framework consists in identifying the set of relevant objects and the set of allowed operations. The whole formalism then relies on the following principle that has clear operational meaning: the resource under consideration cannot be created by allowed operations. Those objects, however, that can be created by allowed operation constitute the free objects of the resource theory.

In the case of entanglement, the relevant objects are quantum states ϱ of an n-partite physical system described by the composite Hilbert space $\mathfrak{H} = \mathfrak{H}_1 \otimes \cdots \otimes \mathfrak{H}_n$. The set of allowed operations is the class of local operations and classical communication (LOCC). An operation from LOCC consists of successive implementation of local operations $\Lambda_1 \otimes \cdots \otimes \Lambda_n : \mathfrak{H} \to \mathfrak{H}$, where each $\Lambda_i : \mathfrak{H}_i \to \mathfrak{H}_i$ is a completely positive map, and communication of the corresponding results among the different parties. Entanglement of a quantum state is then defined as the resource that cannot be created by LOCC. Thus, the free resource in entanglement theory is given by states that can be created by LOCC alone, i.e. by states of the form

$$\varrho = \sum_i p_i \varrho_1^{(i)} \otimes \cdots \otimes \varrho_n^{(i)}, \tag{3.1}$$

where $p_i \geq 0$, $\sum_i p_i = 1$ and $\varrho_j^{(i)}$ quantum states on $\mathfrak{H}_j$ for all i. States of the form of Eq. (3.1) are called separable and it is easy to see that LOCC protocols map separable states into separable states. In turn, states that cannot be created by LOCC are entangled and require a nonlocal quantum resource for their preparation.

The picture becomes more interesting, and more complicated, too, when considering cases where only some of the n parties share entangled states. Consider a partition $\Pi = \{C_1, \ldots, C_k\}$ of $\{1, \ldots, n\}$. A state ϱ is called separable with respect to Π, if it can be written as

$$\varrho = \sum_i p_i \varrho_{C_1}^{(i)} \otimes \cdots \otimes \varrho_{C_k}^{(i)}, \tag{3.2}$$

with probabilities p_i and $\varrho_{C_j}^{(i)}$ a quantum state on $\bigotimes_{r\in C_j} \mathfrak{H}_r$ for all i. Such states are not genuinely n-partite entangled, as they can be created by LOCC with respect to Π, i.e. by local operations of the form $\Lambda_1 \otimes \cdots \otimes \Lambda_k$ with $\Lambda_j : \bigotimes_{r\in C_j} \mathfrak{H}_r \to \bigotimes_{r\in C_j} \mathfrak{H}_r$ and classical communication. In general, one calls a state k-separable if it can be written as the convex sum of states that are separable with respect to some partition of $\{1, \ldots, n\}$ into k groups.

3.2. Operational definition of nonlocality

To define nonlocality of correlations operationally, similar to the case of entanglement, one needs to identify the relevant objects and the set of allowed operations. Then, nonlocality will be defined as the resource that cannot be created using this set of allowed operations alone. The relevant objects are

Resource	Objects	Free Objects	Operations
Entanglement	Quantum states	Separable states	LOCC
Nonlocality	NS boxes	Local correlations	

Table 3.1.: Comparison of entanglement and nonlocality from an operational point of view. The resource of entanglement is defined as the property of quantum states that cannot be created by local operations and classical communication (LOCC). Those states that can be created by LOCC alone, the separable states, then constitute the free resource. To define an analogous framework for the resource of nonlocality one must identify the allowed operations that can be applied to nonsignalling (NS) boxes in a correlation scenario.

clearly nonsignalling boxes P characterised by the joint probability distributions $P(a_1 \dots a_n|x_1 \dots x_n)$. In the operational definition of entanglement we have seen that the allowed operations were given by local operations and classical communication (LOCC), see Table 3.1. What, then, is the analogue of LOCC in the case of nonsignalling boxes?

Let us first look at what corresponds to local operations in the case of boxes. Assume then that the n input-output devices of a given n-partite nonsignalling box are grouped into k groups, so that we can make sense of the term local; of course the case $k = n$ is a valid scenario as well. These k groups should be thought of as k new parties, each of which may act on the devices it has access to. Acting on such input-output devices consists of processing the classical inputs and outputs, i.e. party j may process a given input y_j to determine the input for one of the devices of that group. The obtained output of this first measurement may then be used, together with the provided input y_j, to determine the next measurement choice. Proceeding like this party j will obtain outputs for all its devices and determine its final output from them and the given input y_j. This type of processing is commonly referred to as *wirings*. One can think of the boxes held by one party being wired together in an arbitrary order making use of the previous outputs and the provided input, see Fig. 3.1

More precisely, let $y_1, \dots y_k$ denote the inputs for the wired box. The wiring has to specify how each party j obtains the corresponding output b_j using the input-output devices it has access to. To this end, within every group, an ordering of the devices according to which the group is going to use them needs to be specified. Now, upon receiving y_j group j can use any function f_1^j to compute the input $f_1^j(y_j)$ for the first device yielding an outcome a_1^j;

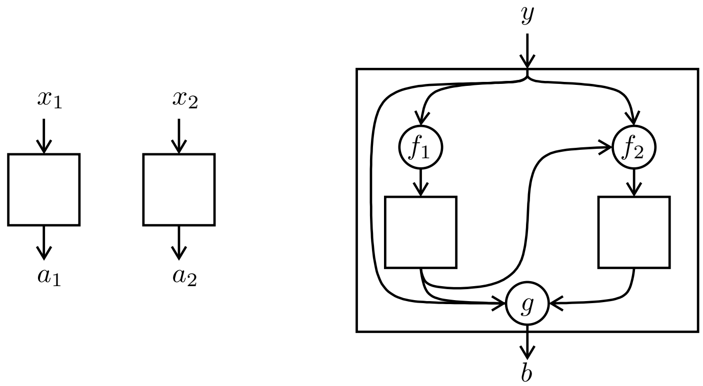


Figure 3.1.: Wiring of a bipartite box to yield a monopartite box. The party that has access to two input-output devices (small boxes on the left) defines a new box (big rectangle on the right) by specifying the order in which the two devices are to be used and functions f_1, f_2, g. If the correlations of the original bipartite box were given by $P(a_1a_2|x_1x_2)$, then the wired box $\tilde{P}$ is characterised by the probabilities $\tilde{P}(b|y) = \sum_{a_1,a_2} P(a_1a_2|f_1(y)f_2(a_1,y))\delta^b_{g(y,a_1,a_2)}$.

to determine the input for the second device a function $f_2^j(y_j, a_1^j)$ is used; in general the p-th input will be determined by $f_p^j(y_j, a_1^j, \dots, a_{p-1}^j)$, where a_p^j is the outcome of the p-th device held by group j. Lastly, the final output b_j of group j is computed by a function $g^j(y_j, a_1^j, \dots, a_p^j)$. We will refer to these actions of the parties, once the inputs are provided, as the *measurement phase*.

One can obtain this general form of a wiring by successive application of the following simpler procedure. Consider the partition of the n parties into one group with access to k devices and $n-k$ groups that hold one device each. As one can always relabel the parties, we assume without loss of generality that the first group holds devices $1, \dots, k$ of the original box. Applying this kind of wiring on the wired box repeatedly all other groupings can be obtained. Thus, we arrive at the following

Definition 1 (Wiring)**.** Let P be a n-partite nonsignalling box. A *wiring* of the first k parties of P in the ordering $1, \dots, k$ is specified by a collection of functions $\{f_i\}_{i=1,\dots,k}$ and a function g. These data define a new m-partite box

P', where $m = n - k + 1$ and the conditional probabilities of P' are given by

$$P'(b_1 \dots b_m | y_1 \dots y_m) = \sum_{\substack{a_1 \dots a_k \\ \text{s.t. } g(y_1, a_1 \dots a_k) = b_1}} P(a_1 \dots a_k b_2 \dots b_m | f_1(y_1) \dots f_k(y_1, a_1 \dots a_{k-1}) y_2 \dots y_m). \tag{3.3}$$

This procedure can be iterated on the resulting box to obtain the general form of a wiring, where the n devices are distributed among s groups. In this case functions like above need to be specified for each group. So, for $1 \leq r \leq s$ one has the the functions $\{f_i^r\}_{i=1,\dots,k_r}$ and g^r, where k_r is the number of devices held by the r-th group.

As mentioned in Chapter 2, we do not specify the input and output alphabets for the different parties. Note, however, that a wiring will in general change these alphabets. For instance, consider for $a_i \in \mathcal{A}_i$ the function $g(a_1, \dots, a_k) = (a_1, \dots, a_k)$ in the above definition. In this case the first output b_1 of the wired box will be an element from $\mathcal{A}_1 \times \dots \times \mathcal{A}_k$. Also the input y_1 may now be from an alphabet different from the alphabet $\mathcal{X}_1$ for the first input of the original box.

It is straightforward to see that wirings of nonsignalling boxes lead to nonsignalling boxes.

Proposition 3.1. *If P' has been obtained from a wiring of the nonsignalling box P, then P' is also nonsignalling.*

Proof. We only have to check that $\sum_{b_1} P'(b_1 \dots b_m | y_1 \dots y_m)$ is independent of y_1 for wired boxes as in Eq. (3.3). So,

$$\begin{aligned}
&\sum_{b_1} P'(b_1 \dots b_m | y_1 \dots y_m) \\
&\quad = \sum_{a_1 \dots a_k} P(a_1 \dots a_k b_2 \dots b_m | f_1(y_1) \dots f_k(y_1, a_1 \dots a_{k-1}) y_2 \dots y_m) \\
&\quad = \sum_{a_1 \dots a_{k-1}} P(a_1 \dots a_{k-1} b_2 \dots b_m | f_1(y_1) \dots f_{k-1}(y_1, a_1 \dots a_{k-2}) y_2 \dots y_m) \\
&\quad = \dots \\
&\quad = P(b_2 \dots b_m | y_2 \dots y_m).
\end{aligned}$$

$\square$

Furthermore, the above definition also covers the case when several nonsignalling boxes are wired to yield a new box. For, given boxes $P_1, \ldots, P_p$, the joint box $P_1 \times \cdots \times P_p$ is a pn-partite nonsignalling box and the above procedure can be applied.

For the operational definition of entanglement classical communication was allowed in addition to local operations. This is justified because one is interested in a property of the state of the physical system according to the formalism of quantum mechanics; classical communication can correlate the locally prepared quantum systems, but only in a classical way. To create entanglement a global preparation or, equivalently, the transmission of quantum systems would be needed.

In the current scenario though, when dealing with nonsignalling boxes, classical communication can only be allowed before the inputs are known. Otherwise the parties could just broadcast their respective input and then decide what to output via classical communication, which would allow them do create arbitrary joint probability distributions. Actually, allowing classical communication after the inputs are known renders the notions local and nonlocal meaningless.

Communication can be permitted, however, if it takes place before the inputs are provided. The parties can use this communication to agree on a certain strategy, e.g. on what wirings they are going to use once they are given the inputs. Another way to prepare the n-partite box before the inputs are provided is possible. Any party may decide to measure one of its devices by choosing any input for that device and announce the measurement outcome together with instructions for the other parties. In function of this result another party may measure one of its systems communicating the obtained outcome and further instructions as well. We will call this procedure of using some of the input-output devices together with classical communication the *preparation phase*.

Formally, these operations can be understood as post-selection, i.e. preparing from the n-partite box P a new box P' by conditioning on a particular outcome given some measurement choice, say, $\tilde{a}_j$ given $\tilde{x}_j$. Formally, we have the

Definition 2 (Post-selection)**.** Let P be an n-partite nonsignalling box. Conditioning P on its j-th outcome to be $\tilde{a}_j$ given that the j-th input is $\tilde{x}_j$ defines the post-selected $(n-1)$-partite box $P_{\tilde{a}_j|\tilde{x}_j}$, characterised by

$$\begin{aligned} &P_{\tilde{a}_j|\tilde{x}_j}(a_1 \ldots a_{j-1} a_{j+1} \ldots a_n | x_1 \ldots x_{j-1} x_{j+1} \ldots x_{n-1}) \\ &\qquad = \frac{1}{P(\tilde{a}_j|\tilde{x}_j)} P(a_1 \ldots a_{j-1} \tilde{a}_j a_{j+1} \ldots a_n | x_1 \ldots x_{j-1} \tilde{x}_j x_{j+1} \ldots x_n). \end{aligned} \tag{3.4}$$

As said before, the obtained outcome can be communicated to the other parties together with further instructions, e.g. concerning which wirings should be used in the measurement phase. Note that in the case when the n devices are originally distributed into k groups, where each group j holds n_j devices, post-selection by group j reduces the number of devices left for the measurement phase. We are now in the position to define the set of allowed operations on a nonsignalling box in a nonlocality scenario.

Definition 3 (Allowed operations)**.** Let P be an n-partite nonsignalling box and $\Pi = \{A_1|\ldots|A_k\}$ a partition of $\{1,\ldots,n\}$. An allowed operation with respect to Π that produces a final k-partite box P_{fin}consists of the following:

(i) Preparation phase: All the cells A_j may perform post-selection on their devices and communicate in several rounds among each other. At the end of this phase they have prepared a m-partite box P' distributed according to a partition $\Pi' = \{B_1|\ldots|B_k\}$, where $B_j \subseteq A_j$ and $m \leq n$.

(ii) Measurement phase: Once the inputs $y_1,\ldots,y_k$ are provided to the cells $B_1,\ldots,B_k$, every cell B_j specifies a wiring for its devices defining the final k-partite box P_{fin} according to Definition 1.

In Section 2.3 we have seen that it is straightforward to generalise the notion of Bell's locality to an n-partite nonsignalling box when considering the partition $\Pi = \{1|\ldots|n\}$. For completeness let us restate this characterisation here.

Definition 4 (Fully local)**.** An n-partite nonsignalling box is said to be *fully local*, if the correlations can be decomposed as

$$P(a_1\ldots a_n|x_1\ldots x_n) = \sum_\lambda p_\lambda P_1^\lambda(a_1|x_1)\cdots P_n^\lambda(a_n|x_n), \tag{3.5}$$

where the P_i^λ are conditional probability distributions and the positive weights p_λ sum to unity.

Now, similar to the case of entanglement, having identified the sets of relevant objects and allowed operations, we want to characterise nonlocality of correlations as the resource that cannot be created by allowed operations. By virtue of Definition 4 we have a notion of nonlocality for the case of the partition $\Pi = \{1|\ldots|n\}$. When considering a general partition of $\{1,\ldots,n\}$ we are led to the following

Definition 5 (Nonlocality of correlations)**.** Let P be an n-partite nonsignalling box and $\Pi = \{A_1, \ldots, A_k\}$ a partition of $\{1, \ldots, n\}$. P is said to be *local with respect to* Π, if every k-partite nonsignalling box P', obtained from P by allowed operations with respect to Π, has a standard local model. Otherwise P is said to be *nonlocal with respect to* Π.

Obviously, this definition of nonlocality is compatible with the standard definition due to Bell, i.e. a fully local n-partite nonsignalling box is local in the sense of Definition 5 with respect to any partition of $\{1, \ldots, n\}$. In particular, choosing the partition $\Pi = \{1|\ldots|n\}$ an n-partite box is local with respect to Π, if and only if it has a standard local model as in Eq. (3.5).

However, similar to the case of entanglement, the situation is more involved when considering partitions of the n parties into $k < n$ groups. The next section discusses what implication our definition of nonlocality has in this case.

3.3. Inconsistencies of standard local models

As already mentioned when describing the operational definition of entanglement, the picture becomes richer when one considers intermediate cases where only some of the n parties share entangled states. Thus, when characterising nonlocality of correlations operationally one must now distinguish not only between local and nonlocal but also take into account these intermediate cases.

From an operational point of view genuine nonlocality of correlations means that for their generation using only allowed operations all the parties must have come together. For definiteness we will consider in what follows the case of three parties A, B, C. The question of genuine nonlocality for this case has previously been studied by Svetlichny (1987). According to him, genuine nonlocality can be characterised by the following

Definition 6 (Svetlichny-bilocal)**.** A tripartite nonsignalling distribution is called *Svetlichny-bilocal*, if it can be decomposed as

$$\begin{aligned} P(abc|xyz) = &\sum_\lambda p_\lambda^A P_A^\lambda(a|x) P_{BC}^\lambda(bc|yz) + \sum_\lambda p_\lambda^B P_B^\lambda(b|y) P_{AC}^\lambda(ac|xz) \\ &+ \sum_\lambda p_\lambda^C P_C^\lambda(c|z) P_{AB}^\lambda(ab|xy), \end{aligned} \tag{3.6}$$

where $0 \leq p_\lambda^A, p_\lambda^B, p_\lambda^C$ and $\sum_\lambda p_\lambda^A + \sum_\lambda p_\lambda^B + \sum_\lambda p_\lambda^C = 1$. Otherwise, the correlations are called *tripartite Svetlichny-nonlocal*.

The above definition can be generalised to an arbitrary number of parties; for our purposes, however, it is sufficient to consider the case of three parties. For definiteness consider only one of the bipartitions, say $A|BC$, and the corresponding decomposition

$$P(abc|xyz) = \sum_{\lambda} p_{\lambda}^{A} P_{A}^{\lambda}(a|x) P_{BC}^{\lambda}(bc|yz). \tag{3.7}$$

It seems justified to call such correlations bilocal with respect to the given partition, as there is a local decomposition for the case when parties B and C are together. Consequently, any valid operation with respect to $A|BC$ in the sense of Definition 3 should map correlations of the form Eq. (3.7) to local correlations. Remarkably, we will see that this is not the case. There are local operations that can map Svetlichny-bilocal correlations to correlations that are nonlocal. This implies that Definition 6 is not compatible with our operational characterisation of nonlocality, as in Definition 5.

Theorem 3.2. *There are a tripartite nonsignalling box P that is Svetlichny-bilocal with respect to the partition $A|BC$ and valid operation with respect to $A|BC$ that takes P to a bipartite nonlocal box P'.*

Proof. We will provide an explicit example. Consider the tripartite nonsignalling correlations with two inputs $x, y, z \in \{0, 1\}$ and two outputs $a, b, c \in \{-1, 1\}$ for each party

$$P(abc|xyz) = \langle\psi|\, M_a^x \otimes N_b^y \otimes O_c^z \,|\psi\rangle\,, \tag{3.8}$$

that can be obtained by local measurements on a pure quantum state $|\psi\rangle$. Explicitly we have

$$\begin{aligned} M_a^0 &= \frac{\mathbb{I}+a\sigma_z}{2} & M_a^1 &= \frac{\mathbb{I}+a\sigma_z}{2} \\ N_b^0 &= \frac{\mathbb{I}+b\sigma_z}{2} & N_b^1 &= \frac{\mathbb{I}+b\sigma_z}{2} \\ O_c^0 &= \frac{\mathbb{I}+c\sigma^+}{2} & O_c^1 &= \frac{\mathbb{I}+c\sigma^-}{2}, \end{aligned} \tag{3.9}$$

where $\sigma^{\pm} = \frac{1}{\sqrt{2}}(\sigma_z \pm \sigma_x)$ and further $|\psi\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)$. These correlations admit a Svetlichny-bilocal decomposition as in Eq. (3.7), as can be confirmed by a linear program. A valid operation for the partition $A|BC$ that can create a nonlocal bipartite box P' can be realised by the simple wiring where C uses as input $f(b) = \frac{1+b}{2}$ with b the output of B, i.e.

$$P'(ac|xy) = \sum_{b} P(abc|xyf(b)). \tag{3.10}$$

To see that this box is nonlocal we calculate the value of the Clause-Horne-Shimony-Holt (CHSH) polynomial

$$\beta(P') = \sum_{x,y} (-1)^{xy} C(x,y), \tag{3.11}$$

where $C(x,y) = \sum_{a,c} ac\ P'(ac|xy)$ to find $\beta(P') = \frac{3}{\sqrt{2}}$ which is greater than the maximal value of 2 for local correlations.

Alternatively, one can use post-selection together with a wiring to obtain an even greater value of the CHSH polynomial. Suppose that B chooses $\tilde{y} = 1$ before the inputs x and z are provided and obtains the outcome b. Then, when the inputs x and z are provided, C uses as input $f(z,b) = bz + \frac{1-b}{2}$, which results in the bipartite correlations

$$P'(ac|xz) = \sum_b P(abc|x\tilde{y}f(z,b)). \tag{3.12}$$

Calculating the CHSH value of P' yields the maximal value $\beta(P') = 2\sqrt{2}$ that can be achieved with quantum systems. □

Let us stress that the above theorem shows that the standard definition of tripartite nonlocality according to Definition 6 is not compatible with the operational definition of nonlocality. The proof of Theorem 3.2 shows that a valid local, i.e. local with respect to $A|BC$, operation can create nonlocality from a box with a decomposition as in Eq. (3.7). Therefore, as nonlocality is the resource that cannot be created from local operations, this decomposition cannot be considered local in the operational sense, even though it seems to provide a local model for the correlations.

To understand how the previous violation of a Bell inequality is possible, it is instructive to have a closer look at the structure of Svetlichny-bilocal decompositions. The distribution P is nonsignalling, which in the end justifies the application of a wiring as done in the proof. However, the individual terms $P_A^\lambda P_{BC}^\lambda$ need not be nonsignalling; the distribution of the variable λ is fine-tuned to yield nonsignalling correlations when taking the average over λ. In particular, a term as P_{BC}^λ may display signalling both from B to C or vice versa for a certain λ, i.e.

$$\sum_b P_{BC}^\lambda(bc|yz) \quad \text{may depend on } y, \tag{3.13}$$

$$\sum_c P_{BC}^\lambda(bc|yz) \quad \text{may depend on } z. \tag{3.14}$$

In this case one cannot interpret λ as the hidden state that would characterise the behaviour of the physical system, as one needs to specify, for a given λ, both y and z to obtain the corresponding outcomes. In a hidden variable model according to Bell, however, knowledge of λ and the input x_i for the i-th party is sufficient to determine the outcome of that party. Thus, a decomposition including such terms cannot be considered the adequate physical model for a situation where one of the parties measures first.

Despite this physical argument against Svetlichny-bilocal decompositions, one is tempted to think that formally applying a wiring between B and C to every term in the decompositions should lead to a bipartite local model. However, this is in general not the case.

Proposition 3.3. *Given conditional probabilities $P^\lambda(bc|yz)$ with signalling from C to B, there is a wiring from B to C such that the resulting object is not a conditional probability distribution.*

Proof. Signalling from C to B means that there are b_0, y_0, z_1, z_2 such that

$$\sum_c P^\lambda(b_0c|y_0z_1) \neq \sum_c P^\lambda(b_0c|y_0z_2). \tag{3.15}$$

Now define the wired object $\tilde{P}$ through

$$\tilde{P}(c|y) = \sum_b P^\lambda(bc|yf(b)), \tag{3.16}$$

with

$$f(b) = \begin{cases} z_2 & \text{if } b = b_0 \\ z_1 & \text{otherwise.} \end{cases} \tag{3.17}$$

Next calculate

$$\begin{aligned}\sum_c \tilde{P}(c|y_0) &= \sum_c P^\lambda(b_0c|y_0z_2) + \sum_{b\neq b_0}\sum_c P^\lambda(bc|y_0z_1) \\ &\neq \sum_c P^\lambda(b_0c|y_0z_1) + \sum_{b\neq b_0}\sum_c P^\lambda(bc|y_0z_1) \\ &= 1\end{aligned}$$

to conclude that $\tilde{P}$ is not a conditional probability. □

This shows that the local decomposition of a Svetlichny-bilocal box can in general not be used to construct a local model for the wired box. Correlations as in Theorem 3.2 that allow for the creation of nonlocality by local operations must therefore involve terms in their decomposition that display signalling. Whether the converse is true, i.e. whether correlations for which every Svetlichny-bilocal decomposition contains signalling terms can be mapped by local operations to nonlocal bipartite correlations, remains an open question.

As noted before, the definition of Svetlichny-bilocal can be generalised to n-partite nonsignalling distributions to yield notions of k-locality. For definiteness this section considered the case of three parties and bilocal decompositions, as the aim was to show the inconsistency of such decompositions with our operational definition of nonlocality.

3.4. Time-ordered local models

When considering n distant parties we have seen that our operational definition leads to the standard definition of locality due to Bell as in Definition 4. In this case the existence of a local hidden variable model for the correlations is equivalent to their being local in the operational sense. However, the analysis of Svetlichny-bilocal decompositions showed, that these decompositions cannot be considered local within the current operational framework. Therefore, the question naturally arises as to whether one can find forms of bilocality, or more generally, k-locality that would be consistent with the operational definition of locality. In other words, of what form must k-local models be to capture the notion of locality? Of particular interest will be the case of bilocality, i.e. $k = 2$, as this will allow the definition of genuine nonlocality as those correlations that are not bilocal.

The previous section also showed that the trouble with Svetlichny's bilocal models was the appearance of possibly two-way signalling terms in the decomposition. These terms in general lead to the mentioned inconsistencies when considering post-selection or wirings. From this it is clear that a possible way to avoid such inconsistencies is to demand that a local model be a decomposition into nonsignalling distributions only. Thus, considering for simplicity the tripartite case and only one of the partitions, one can define models of the form

$$P(abc|xyz) = \sum_{\lambda} p_{\lambda}^{A} P_{A}^{\lambda}(a|x) P_{BC}^{\lambda}(bc|yz), \tag{3.18}$$

where P_{BC}^{λ} is nonsignalling for all λ. One can operationally understand these correlations as correlations obtained from collaborating parties B and C that share nonsignalling resources, where BC is spatially separated from A. This definition is, however, restrictive as it does not allow for signalling, i.e. communication, between the collaborating parties. If one thinks of B and C as two individual parties that at the same time are close, i.e. not spatially separated, it is difficult to motivate why the no-signalling principle should hold for them. How can one then incorporate signalling into local models without being led into the inconsistencies encountered in the analysis of the Svetlichny-bilocal decompositions? The answer we propose here consists of models that allow for signalling among collaborating parties only in one direction, which corresponds to the temporal order in which the parties measure their systems. As we do not want to assume that the order is known in advance, we must demand that there exists such a decomposition for every possible ordering among the collaborating parties. This leads us to the following

Definition 7 (Time-ordered local models)**.** Let P be an n-partite nonsignalling box and $\Pi = \{C_1, \dots, C_k\}$ a partition of $\{1, \dots, n\}$. P is said to admit a *time-ordered local model with respect to* Π, if for every collection of permutations $(\sigma_1, \dots, \sigma_k) \in S_{|C_1|} \times \dots \times S_{|C_k|}$ it can be decomposed as

$$P = \sum_\lambda p_\lambda P^\lambda_{\sigma_1} \dots P^\lambda_{\sigma_k}, \tag{3.19}$$

where each of the conditional probabilities $P^\lambda_{\sigma_j}$ is defined for the $r_j = |C_j|$ parties of the j-th cell of the partition and satisfies

$$\sum_{a_{\sigma_j(m+1)} \dots a_{\sigma_j(r_j)}} P^\lambda_{\sigma_j}(a_1 \dots a_{r_j} | x_1 \dots x_{r_j}) \\ = P^\lambda_{\sigma_j}(a_{\sigma_j(1)} \dots a_{\sigma_j(m)} | x_{\sigma_j(1)} \dots x_{\sigma_j(m)}) \tag{3.20}$$

for all $1 \leq m < r_j$.

If the conditional probabilities $P^\lambda_{\sigma_i}$ are all nonsignalling, i.e. $P^\lambda_{\sigma_i} = P^\lambda_{\mathrm{id}}$ for all $\sigma_i \in S_{r_i}$ and for all $1 \leq i \leq n$, then P is said to have a *nonsignalling local model with respect to* Π.

So, for a fixed partition Π, a given λ, and any ordering of the r_j parties within each cell C_j of the partition, there is a conditional probability distribution that in general is signalling, but allows only for signalling in one direction determined by the ordering σ_j. Operationally, this means that within every cell of the partition the temporal ordering according to which the system of that cell is measured can be chosen independently of the ordering within the other cells.

The main result of this chapter is to show that this definition provides local models that are compatible with our operational definition of nonlocality.

Theorem 3.4. *Time-ordered local models are compatible with the operational definition of nonlocality. Formally, let $P^{(i)}$ be an n-partite nonsignalling box with a time-ordered local model with respect to $\Pi_i = \{C_1^i, \ldots, C_{r_i}^i\}$, a partition of $\{(i-1)n+1, \ldots, in\}$, for $1 \leq i \leq k$. Define the product box $P = P^{(1)} \times \ldots \times P^{(k)}$ and the product partition $\Pi_{\text{prod}} = \bigcup_i \Pi_i$. Then for any partition $\Pi = \{C_1, \ldots, C_r\}$ of $\{1, \ldots, kn\}$ coarser than Π_{prod} the following holds:*

(i) P is time-ordered local with respect to Π.

(ii) Any post-selection within C_j takes P to a box P' that has a time-ordered local model with respect to $\Pi' = \{C_1, \ldots, C_{j-1}, C', C_{j+1}, \ldots, C_r\}$ with $C' \subset C_j$.

(iii) Any wiring within C_j maps P to a box P' that has a time-ordered local model with respect to $\Pi' = \{C_1, \ldots, C_{j-1}, C', C_{j+1}, \ldots, C_r\}$ with $|C'| = 1$.

The above result has a clear operational meaning. Given an arbitrary number of boxes, where each box admits a time-ordered local model with respect to some partition, one distributes these boxes among r parties. Each of the r parties can hold several cells of one or several boxes, but one cell of a box cannot be shared by two or more parties (that is the notion of a partition coarser than the product one). Applying allowed operations, i.e. post-selection and local wirings, the parties will end up with a box that is again time-ordered local, see Fig. 3.2 for an example of two tripartite boxes. In particular, they cannot create any nonlocality with respect to the partition according to which the boxes were distributed among them, contrary to what happened in the case of Svetlichny-bilocal decompositions.

Proof. It is clear that P is time-ordered local with respect to the product partition. Every coarser partition can be obtained by successively joining cells of the product partition; for a cell obtained from joining C_i and C_j the corresponding conditional probability $P^{\lambda}_{C_i \cup C_j, \sigma}$ will be given by the product $P^{\lambda}_{C_i,\sigma_i} P^{\lambda}_{C_j,\sigma_j}$, where for a given permutation σ of the elements of $C_i \cup C_j$ one has to choose the permutations σ_i, σ_j as the permutations induced on C_i and C_j by σ respectively. This product clearly fulfils the condition Eq. (3.20) for time-ordered local models. This shows (i).

To see (ii), consider a post-selection in the j-th cell C_j and denote the outputs and inputs belonging to C_j as $a_1 \ldots a_{r_j}$ and $x_1 \ldots x_{r_j}$ respectively. Let the post-selection be on outcome $a_p = \tilde{a}$ given the setting $x_p = \tilde{x}$. We want to show that the post-selected box $P'_{\tilde{a}|\tilde{x}}$ can be decomposed as

$$P'_{\tilde{a}|\tilde{x}} = \sum_{\lambda} \tilde{p}_{\lambda} P^{\lambda}_{\sigma_1} \ldots P^{\lambda}_{\sigma_{j-1}} P'^{\lambda}_{\tau} P^{\lambda}_{\sigma_{j+1}} \ldots P^{\lambda}_{\sigma_r}, \tag{3.21}$$

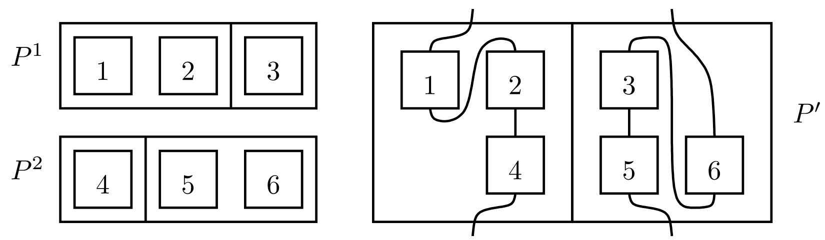


Figure 3.2.: Compatibility of time-ordered local models with the operational definition of locality. Starting from two tripartite boxes P^1 and P^2 that have time-ordered local models with respect to the partitions $\Pi_1 = \{1,2|3\}$ and $\Pi_2 = \{4|5,6\}$, one considers the product box $P^1 \times P^2$. This product box admits a time-ordered local model with respect to the product partition $\{1,2|3|4|5,6\}$ and to any partition coarser than the product partition, as e.g. the bipartition $\Pi = \{1,2,4|3,5,6\}$. If the two parties corresponding to Π now apply allowed operations with respect to Π, they end up with the bipartite local box P'.

for arbitrary permutations $\sigma_i \in S_{r_i}$ and $\tau \in S_{r_j-1}$. Think of τ as a permutation that permutes the elements $\{1,\ldots,p-1,p+1,\ldots,r_j\}$, and define the permutation $\sigma \in S_{r_j}$ by

$$\sigma(i) = \begin{cases} p & \text{if } i = 1, \\ \tau(i-1) & \text{if } 2 \leq i \leq p, \\ \tau(i) & \text{if } p < i \leq r_j. \end{cases} \tag{3.22}$$

This permutation corresponds to the ordering that starts with p followed by the order of the remaining indices as specified by τ. With this σ and choosing $m = 1$ in Eq. (3.20) we get

$$P^\lambda_\sigma(\tilde{a}|\tilde{x}) = \sum_{a_{\sigma(2)}\ldots a_{\sigma(r_j)}} P^\lambda_\sigma(a_1\ldots a_{p-1}\tilde{a}a_{p+1}\ldots a_{r_j}|x_1\ldots x_{p-1}\tilde{x}x_{p+1}\ldots x_{r_j}). \tag{3.23}$$

With this we can define a time-ordered model for the post-selected box by

$$P'_{\tilde{a}|\tilde{x}} = \sum_\lambda \tilde{p}_\lambda P^\lambda_{\sigma_1}\ldots P^\lambda_{\sigma_{j-1}} P'^\lambda_\tau P^\lambda_{\sigma_{j+1}}\ldots P^\lambda_{\sigma_r}, \tag{3.24}$$

where the weights are given by

$$\tilde{p}_\lambda = p_\lambda \frac{P^\lambda_\sigma(\tilde{a}|\tilde{x})}{P(\tilde{a}|\tilde{x})} \tag{3.25}$$

and the conditional probability for the now $r_j - 1$ parties of the j-th cell is given by

$$\begin{aligned} &P'^{\lambda}_{\tau}(a_1 \dots a_{p-1} a_{p+1} \dots a_{r_j} | x_1 \dots x_{p-1} x_{p+1} \dots x_{r_j}) \\ &\qquad = \frac{1}{P^{\lambda}_{\sigma}(\tilde{a}|\tilde{x})} P^{\lambda}_{\sigma}(a_1 \dots a_{p-1} \tilde{a} a_{p+1} \dots a_{r_j} | x_1 \dots x_{p-1} \tilde{x} x_{p+1} \dots x_{r_j}). \end{aligned} \tag{3.26}$$

Assertion (ii) now follows from the properties of P^{λ}_{σ}.

To show (iii), consider the first cell C_1, where we assume without loss of generality that the devices are used in the order $1, \dots, r_1$. Now, write $t = r_1$ and let $s = nk - t + 1$. After the wiring we have

$$\begin{aligned} &P'(b_1 \dots b_s | y_1 \dots y_s) \\ &\quad = \sum_{\substack{a_1 \dots a_t \\ \text{s.t. } g(y_1 a_1 \dots a_t) = b_1}} P(a_1 \dots a_t b_2 \dots b_s | f_1(y_1) \dots f_t(y_1, a_1 \dots a_{t-1}) y_2 \dots y_s). \end{aligned} \tag{3.27}$$

This can clearly be written as

$$P' = \sum_{\lambda} p_{\lambda} \tilde{P}^{\lambda} P^{\lambda}_{\sigma_2} \dots P^{\lambda}_{\sigma_r}, \tag{3.28}$$

where $\tilde{P}^{\lambda}$ is given by

$$\tilde{P}^{\lambda}(b_1|y_1) = \sum_{a_1 \dots a_t} P^{\lambda}_{\mathrm{id}}(a_1 \dots a_t | f_1(y_1) \dots f_t(y_1, a_1 \dots a_{t-1})) \delta^{b_1}_{g(y_1 a_1 \dots a_t)}. \tag{3.29}$$

Condition Eq. (3.20) ensures that $\tilde{P}^{\lambda}$ is a well-defined conditional probability, as

$$\begin{aligned} \sum_{b_1} \tilde{P}^{\lambda}(b_1|y_1) &= \sum_{a_1 \dots a_t} P^{\lambda}_{\mathrm{id}}(a_1 \dots a_t | f_1(y_1) \dots f_t(y_1, a_1 \dots a_{t-1})) \\ &= \sum_{a_1 \dots a_{t-1}} P^{\lambda}_{\mathrm{id}}(a_1 \dots a_{t-1} | f_1(y_1) \dots f_{t-1}(y_1, a_1 \dots a_{t-2})) \\ &= \dots \\ &= \sum_{a_1} P^{\lambda}_{\mathrm{id}}(a_1 | f_1(y_1)) = 1. \end{aligned}$$

The wirings within the other cells can be treated analogously. □

As mentioned before, time-ordered local models include by definition also non-signalling local models. Interestingly one can show that this inclusion is in general strict.

Proposition 3.5. *Let $\Pi = \{C_1, \ldots, C_k\}$ be a partition of $\{1, \ldots, n\}$ and let* TOL *and* NSL *denote the set of n-partite nonsignalling boxes that allow for time-ordered local models and nonsignalling local models with respect to Π respectively. Then* $\mathsf{TOL} \supsetneq \mathsf{NSL}$*, unless $k = n$, in which case* $\mathsf{TOL} = \mathsf{NSL}$.

Proof. Obviously, $k = n$ means that $|C_i| = 1$ for all i, in which case both the notion of time-ordered local and nonsignalling local reduce to the standard notion of locality as in Definition 4. Let us first show the assertion for the case of the partition $\{1|2,3\}$, from which the general case will follow. Now, let $P \in \mathsf{NSL}$, i.e. $P(abc|xyz) = \sum p_\lambda P_A^\lambda(a|x) P_{BC}^\lambda(bc|yz)$ with P_{BC}^λ nonsignalling for all λ, and consider the following expression for a tripartite box

$$\beta(P) = P(000|000) + P(110|011) + P(011|101) + P(101|110), \tag{3.30}$$

know as "Guess Your Neighbour's Input" (GYNI) (Almeida et al., 2010). Without loss of generality we can assume the functions P_A^λ to be deterministic so that we get

$$\beta(P) \leq \sum_\lambda p_\lambda \begin{cases} P_{BC}^\lambda(00|00) + P_{BC}^\lambda(11|01) \\ P_{BC}^\lambda(00|00) + P_{BC}^\lambda(01|10) \\ P_{BC}^\lambda(10|11) + P_{BC}^\lambda(11|01) \\ P_{BC}^\lambda(10|11) + P_{BC}^\lambda(01|10). \end{cases} \tag{3.31}$$

As P_{BC}^λ is nonsignalling this expression can be bounded as follows

$$\begin{aligned} \beta(P) &\leq \sum_\lambda p_\lambda \begin{cases} P_B^\lambda(0|0) + P_B^\lambda(1|0) \\ P_C^\lambda(0|0) + P_C^\lambda(1|0) \\ P_C^\lambda(0|1) + P_C^\lambda(1|1) \\ P_B^\lambda(1|1) + P_B^\lambda(0|1) \end{cases} && (3.32) \\ &\leq 1. && (3.33) \end{aligned}$$

However, there are tripartite nonsignalling boxes from TOL that can attain values greater than unity for the GYNI expression; in particular, Appendix A contains a tripartite time-ordered local distribution $P \in \mathsf{TOL}$ with $\beta(P) = 7/6$. This shows that in the case of three parties $\mathsf{NSL} \subsetneq \mathsf{TOL}$. Using this argument, the general case now follows by considering tripartite marginal distribution of the n-partite boxes with two parties from one cell and the third from another. $\square$

Thus, the set of correlations admitting a nonsignalling local model NSL constitutes a set of correlations compatible with our operational definition of nonlocality, but it is not the largest set with this property. Whether the set TOL is the largest such set, on the other hand, remains an open question. Given the clear operational definition of time-ordered local models, we conjecture this to be the case, but a general proof is missing. Even a proof of this conjecture for small number of parties and small input and output alphabets is difficult, as it is hard to characterise the extreme points of TOL. This has to do with the conditional probabilities $P^{\lambda}_{\sigma_i}$ constituting a time-ordered local model; they are themselves not extreme points of the polytope TOL due to the signalling they display in general.

To conclude, we have introduced a novel framework for the characterization of nonlocality which has an operational motivation and captures the role of nonlocality as a resource for device-independent quantum information processing. In spite of its simplicity, the framework questions the current understanding of genuine multipartite nonlocality, as the standard definition adopted by the community is inconsistent with it. Similar conclusions are reached from another perspective by Barrett et al. (2011).

By introducing time-ordered local models we provided an alternative where consistency with the operational definition of nonlocality is recovered. As mentioned in the discussion of Svetlichny-bilocal decompositions, the main open question is whether time-ordered local models constitute the largest set compatible with the allowed operations. We have seen that two-way signalling terms in such decompositions can lead to inconsistencies, but it is not known whether this is always the case, i.e. whether nonsignalling boxes that do not allow for a time-ordered local model can always be mapped by allowed operations to nonlocal correlations. However, as time-ordered models have such a clear operational meaning and incorporate signalling in a way that corresponds to the physical situation considered, the author of this work conjectures that TOL actually constitutes the largest set compatible with the operational definition of nonlocality. In particular, this would imply that a box is genuine n-partite nonlocal, if and only if it cannot be written as a convex sum of time-ordered bilocal models.

4. Quantum correlations require multipartite information principles

Quantum mechanics exhibits many characteristic properties that distinguish it from classical physics. For instance, quantum states cannot be cloned, quantum mechanics only predicts probabilities for measurement outcomes, and local measurements on quantum states can give rise to correlations that are stronger than any classical correlations. These traits are, however, not unique to quantum mechanics. In the framework of generalised probabilistic theories one can study classes of theories, including quantum mechanics as a special case, that share these phenomena.

One approach to characterise quantum mechanics focuses on the correlations among distant observers that are possible within the theory. Rather than reconstructing the entire formalism of states, transformations and measurements, one tries to identify a general principle, formulated only in terms of correlations, that would allow to single out the set of quantum correlations. Recently, information theoretic concepts have been advocated as a possibility to achieve this task (van Dam, 2000; Clifton et al., 2003).

This chapter shows a fundamental limitation of this approach: no principle based on bipartite information concepts is able to single out the set of quantum correlations for an arbitrary number of parties. Our results reflect the intricate structure of quantum correlations and imply that new and intrinsically multipartite information concepts are needed for their full understanding.

4.1. Nonsignalling versus quantum correlations

Let us look back at the definition of nonsignalling and quantum correlations to see why one would want to have an information principle for quantum correlations in the first place.

Nonsignalling correlations First, consider the situation in which the measurements by the observers define space-like separated events, i.e. during the process starting with the measurement choice and ending with recording the outcome no signal can travel to any of the other parties. Under these conditions, the laws of special relativity guarantee that the statistics seen by any subset of k observers are independent of the measurement choices of the remaining $n-k$ parties.

Indeed, if this was not the case, the $n-k$ observers could use this dependence to signal to the first k parties, even though they are not causally connected. Mathematically, the impossibility of faster-than-light communication is imposed on the conditional probabilities by requiring that for all $1 \leq j \leq n$

$$\sum_{a_j} P(a_1 \dots a_n | x_1 \dots x_n) \tag{4.1}$$

be independent of x_j. In other words, the marginal conditional probability distributions are well defined. These linear constraints define the convex polytope of nonsignalling correlations.

A subset of the nonsignalling correlations is the set of correlations P_{L} having a local hidden-variable model, i.e. P_{L} can be decomposed as

$$P_{\mathrm{L}}(a_1 \dots a_n | x_1 \dots x_n) = \sum_{\lambda} p_\lambda P_1^\lambda(a_1|x_1) \dots P_n^\lambda(a_n|x_n), \tag{4.2}$$

where $p_\lambda \geq 0$ with $\sum_\lambda p_\lambda = 1$ and P_i^λ are conditional probability distributions for all $1 \leq i \leq n$ and λ. These correlations are also called local or classical and have a clear operational meaning: they can be established among the observers when each of them produces locally the outcome a_i using the input x_i and some pre-established classical instructions, denoted by λ. As first shown by Bell (1964), they satisfy some non-trivial linear constraints, known as Bell inequalities. It can also be shown that some correlations are local if, and only if, they are compatible with the no-signalling principle and determinism (Valentini, 2002). Indeed, they can always be decomposed into mixtures with the functions P_i^λ being deterministic (Fine, 1982).

As we see, both nonsignalling and classical correlations can be defined by appealing to clear physical principles solely based on correlations. Mathematically, the defining conditions take the form of linear constraints on the conditional probability distribution P.

Quantum correlations Assuming that the parties share a quantum state and perform measurements as described by the rules of quantum mechanics, the resulting correlations are called quantum. Formally, one assigns a (finite-dimensional) Hilbert space $\mathbb{C}^{d_i}$ to every party, yielding the total Hilbert space of the system $\mathfrak{H} = \bigotimes_i \mathbb{C}^{d_i}$. The state of the system is then a density matrix

ϱ acting on $\mathfrak{H}$, i.e. a linear map $\varrho : \mathfrak{H} \to \mathfrak{H}$ with $\varrho \succeq 0$ and $\operatorname{tr} \varrho = 1$. For every party i each measurement x_i is described by a positive-operator valued measure (POVM), i.e. a collection of positive operators $\{F^i_{a_i|x_i}\}$ on $\mathbb{C}^{d_i}$, where

$$\sum_{a_i} F^i_{a_i|x_i} = \mathbb{I}_{d_i} \tag{4.3}$$

for every x_i. The observed correlations are then given by the Born rule, so that the conditional probabilities read

$$P(a_1 \dots a_n | x_1 \dots x_n) = \operatorname{tr}(\varrho\, F^1_{a_1|x_1} \otimes \dots \otimes F^n_{a_n|x_n}). \tag{4.4}$$

Quantum correlations are known to lie between the set of classical and general nonsignalling correlations as there exist quantum correlations that violate a Bell inequality and therefore have no classical analogue (Bell, 1964), and nonsignalling correlations that are not quantum (Popescu and Rohrlich, 1994), i.e., they cannot be written in the form Eq. (4.4). Contrary to the definition of nonsignalling and classical correlations, the set of quantum correlations is not characterised in terms of general principles but is defined by making explicit reference to the mathematical structure of quantum mechanics. Thus, despite having a clear mathematical definition as in Eq. (4.4), the set of quantum correlations lacks a natural interpretation in terms of physical principles. As mentioned earlier, it has been suggested that information theoretic concepts could provide such principles for quantum correlations.

4.2. Information principles

This section makes more precise what is usually meant by an information principle for quantum correlations. In general, one considers a correlation scenario where several parties have access to some physical resource state and possibly access to additional information resources such as classical communication. The formulation of these scenarios often take the form of an information processing task. Relating an information theoretic quantity of the correlations and how well the parties can achieve the given task constitutes an information principle.

To arrive at a principle that singles out quantum correlations within the set of nonsignalling correlations, one would then need to show that correlations satisfy the principle, if and only if they are quantum. Proving that quantum correlations do fulfil the principle is typically the easy part of this task; it is the only-if part that makes this problem hard. Say one found a nonsignalling probability distribution P known to be not quantum that fulfilled the principle. Wouldn't this prove the principle too weak to single out quantum correlations? Not quite. It might well be that P satisfies the principle, but several copies of P do not. In this case one would have to consider P as non-physical on the grounds of the given principle, as copies of P give rise to correlations that violate the principle. Thus, one rather needs to look at sets S of nonsignalling boxes which have the property that the application of valid operations within the considered scenario to an arbitrary number of members from S always leads to correlations that satisfy the principle. Proving that a given principle actually singles out quantum correlations, then corresponds to showing that the largest set S that fulfils the principle is the set of quantum correlations.

It is difficult to give a precise mathematical definition of an information principle in general terms. Therefore, we illustrate the idea behind this approach by presenting two prominent examples of information principles that were proposed as candidates to distinguish between general nonsignalling correlations and the actual observed physical correlations.

One of the first to show how access to stronger than quantum correlations leads to implausible consequences was van Dam (2000). He considered the consequences of arbitrary nonsignalling correlations for the concept of *communication complexity*, a notion introduced by Yao (1979) that studies the following problem. Two parties A and B are given an n-bit string x and y respectively, and the task is for one of them, say B, to compute the value $f(x, y)$ of a given function f with as little communication from A to B as possible. Clearly, the task can always be achieved if A sends the entire string x, i.e. a total of n bits, to B. Communication complexity is said to become *trivial*, if this task can be achieved for any function with a constant, i.e. independent of n, amount of communication. Van Dam (2000) showed that the parties can in fact achieve this task for any boolean function $f : \{0, 1\}^n \times \{0, 1\}^n \to \{0, 1\}$ by communication of just one bit, if they share n copies of the PR-box

$$P_{\mathrm{PR}}(ab|xy) = \begin{cases} \frac{1}{2} & a + b \equiv xy \pmod 2 \\ 0 & \text{otherwise}, \end{cases} \tag{4.5}$$

thereby making communication complexity trivial. Thus, formulated as a principle, one would say that the only allowed correlations in nature are those that do not make communication complexity trivial. These results were generalised (Brassard et al., 2006; Brunner and Skrzypczyk, 2009) to exclude not just the extremal PR-box but further post-quantum correlations from the set of nonsignalling correlations as non-physical, as they would imply trivial communication complexity. Whether this principle can single out quantum correlations is, however, unknown.

The principle of *information causality*, introduced by Pawlowski et al. (2009), can be understood as a generalisation of the no-signalling principle. It considers a scenario of two parties A and B sharing some physical resources; the first party A further holds a data set unknown to B. As a principle, information causality states that the information B can gain of the data set using its local resources and a classical message of m bits received by A is at most m bits. In the case that A sends no information to B, i.e. $m = 0$, information causality is just the no-signalling principle. Pawlowski et al. (2009) showed that quantum correlations, and therefore also classical correlations, fulfil information causality, whereas the principle can be violated by some nonsignalling correlations. In particular, if the data set of A consists just of two bits $\{a_0, a_1\}$ and the two parties share a PR-box, then a message of $m = 1$ bit enables B to learn any of the two bits held by A perfectly; B cannot learn both bits, but by choosing its local measurements B can decide which bit to know perfectly. As said, this is impossible, both classically and within quantum mechanics. In fact, the optimal strategy in this case consists of A sending one of its bits to B. Information causality further recovers Tsirelson's bound, the bound for the strongest correlations allowed by quantum mechanics in the CHSH scenario, as correlations stronger than Tsirelson's bound always violate the principle (Pawlowski et al., 2009). This shows that access to arbitrary nonsignalling correlations lead to implausible consequences from a information-theoretic point of view. However, this is not enough to single out the entire set of quantum correlations, as there are also nonsignalling correlations outside the quantum set that are weaker than Tsirelson's bound.

Subsequent work by Allcock et al. (2009) could exclude further bipartite correlations outside the quantum set as non-physical assuming the validity of information causality. The authors showed that information causality recovers parts of the boundary of the quantum set, thereby providing further evidence that the principle might single out quantum correlations. The principle was also successfully applied to other scenarios, such as Hardy's nonlocality (Ahanj et al., 2010) and macroscopic locality (Cavalcanti et al., 2010). However, the question as to whether information causality is sufficient to completely characterise the set of quantum correlations remains open.

As information causality is defined in a bipartite scenario (Pawlowski et al., 2009), it is not immediately clear how the principle should be applied to the case of more than two parties. The fact, however, that the principle is a generalisation of the no-signalling principle suggests that it should be applied to the multipartite case the same way as no-signalling is applied to this case. The no-signalling principle for n parties is just the bipartite no-signalling principle applied to every bipartition of the n-partite system. Therefore, the natural generalisation of the principle of information causality to n parties reads: an n-partite box P fulfils information causality in the multipartite scenario, if P fulfils information causality for every bipartition of the system. This generalisation of information causality has recently been applied to the study of extremal tripartite non-signalling correlations (Yang et al., 2012).

For the multipartite generalisation of communication complexity one can consider k parties, where each party i holds a bit string x_i of length n for $1 \leq i \leq k$. Communication complexity is trivial in this scenario, if there is a communication protocol with a constant C such that every party needs to broadcast at most C bits to all other parties in order to evaluate any given function $f : \{0,1\}^{nk} \to \{0,1\}$ at $(x_1, \ldots, x_k)$. So, trivial multipartite communication complexity implies trivial communication complexity for every bipartition of the system with communication of at most $(k-1)C$ bits. Hence, the principle stating that the allowed multipartite correlations are those that do not make multipartite communication complexity trivial can be imposed by demanding communication complexity to be non-trivial for every bipartition of the multipartite system.

These observations show that in many instances information principles are essentially bipartite, as their formulation for the multipartite case consists in the application of a bipartite principle to every bipartition. This motivates the following

Definition 8 (Essentially bipartite principles)**.** An information principle IP is called *essentially bipartite*, if the following holds for all n-partite nonsignalling boxes P:

$$P \text{ fulfils IP for all bipartitions of } \{1, \ldots, n\} \;\Leftrightarrow P \text{ fulfils IP} \tag{4.6}$$

4.3. Insufficiency of bipartite principles

In the following we show that any physical principle that, similarly to the no-signalling principle, is essentially bipartite is not sufficient to characterise the set of quantum correlations. We show this by finding correlations that, on one hand, fulfil any information principle based on bipartite concepts and, on the other hand, are outside the set of quantum correlations. To ensure that our distributions are compatible with any bipartite information principle that aims to single out quantum correlations, we will define a set of correlations that behave classically in any bipartite sense.

To this end we make use of the time-ordered local models defined in Chapter 3. We will demand the nonsignalling boxes correlations to have a time-ordered bilocal model for all bipartitions of the system at the same time.

Definition 9 (Fully time-ordered bilocal)**.** An n-partite nonsignalling distribution P is said to admit a *fully time-ordered bilocal model*, if it admits a time-ordered local model with respect to every bipartition $\{C_1|C_2\}$ of $\{1,\ldots,n\}$. More explicitly, for all $\sigma_1 \in S_{C_1}$,$\sigma_2 \in S_{C_2}$ the box P can be decomposed as

$$P = \sum_{\lambda} p_{\lambda}^{C_1|C_2} P_{C_1,\sigma_1}^{\lambda} P_{C_2,\sigma_2}^{\lambda}, \tag{4.7}$$

where for $i = 1,2$ the conditional probabilities $P_{C_i,\sigma_i}^{\lambda}$ are defined for the $|C_i| = r_i$ parties of C_i and fulfil

$$\begin{aligned}\sum_{a_{\sigma_i(m+1)},\ldots,a_{\sigma_i(r_i)}} & P_{C_i,\sigma_i}^{\lambda}(a_1 \ldots a_{r_i}|x_1 \ldots x_{r_i}) \\ &= P_{C_i,\sigma_i}^{\lambda}(a_{\sigma_i(1)} \ldots a_{\sigma_i(m)}|x_{\sigma_i(1)} \ldots x_{\sigma_i(m)})\end{aligned} \tag{4.8}$$

for all $1 \leq m \leq r_i - 1$.

In the case of only three parties it is instructive to write the above definition more explicitly: a tripartite nonsignalling distribution P has as fully time-ordered bilocal model, if

$$\begin{aligned}P(a_1a_2a_3|x_1x_2x_3) &= \sum_{\lambda} p_{\lambda}^{i|jk} P_i^{\lambda}(a_i|x_i) P_{j\to k}^{\lambda}(a_ja_k|x_jx_k) \\ &= \sum_{\lambda} p_{\lambda}^{i|jk} P_i^{\lambda}(a_i|x_i,) P_{j\leftarrow k}^{\lambda}(a_ja_k|x_jx_k)\end{aligned} \tag{4.9}$$

for $(i,j,k) = (1,2,3),(2,3,1),(3,1,2)$, with the distributions $P_{j\to k}^{\lambda}$ and $P_{j\leftarrow k}^{\lambda}$ obeying the conditions

$$\sum_{a_k} P_{j\to k}^{\lambda}(a_ja_k|x_jx_k) = P_{j\to k}^{\lambda}(a_j|x_j), \tag{4.10}$$

$$\sum_{a_j} P_{j\leftarrow k}^{\lambda}(a_ja_k|x_jx_k) = P_{j\leftarrow k}^{\lambda}(a_k|x_k). \tag{4.11}$$

As explained in Chapter 3 the operational meaning of these models can be understood in the following way. Consider the bipartition 1|23 for which systems 2 and 3 form one cell of the partition. In this situation, the correlations $P(a_1a_2a_3|x_1x_2x_3)$ can be simulated by a classical random variable λ distributed according to $p_\lambda^{1|23}$ and shared between 1 and 23. Given λ, 1 generates its output according to the distribution $P_1^\lambda(a_1|x_1)$, whereas, and depending on which of the parties 2 and 3 measures first, 2 and 3 use either $P_{2\to3}^\lambda(a_2a_3|x_2x_3)$ or $P_{2\leftarrow3}^\lambda(a_2a_3|x_2x_3)$ to produce the two measurement outcomes. Likewise, any other bipartition of systems $1,2,3$ admits a classical simulation.

By construction, the set of tripartite fully time-ordered models is convex and includes all tripartite probability distributions of the form of Eq. (4.2). Furthermore, as a consequence of Theorem 3.4 from Chapter 3 we have that all bipartite correlations that can be created by allowed operations from an arbitrary number of boxes that are fully time-ordered bilocal are classical. Therefore, the set of fully time-ordered bilocal correlations constitute a set that satisfies every essentially bipartite information principle.

Proposition 4.1. *If $P_1,\ldots,P_k$ are n-partite nonsignalling boxes, each of which admits a fully time-ordered local model, then all bipartite correlations that can be obtained from them are local.*

Proof. Just observe that the product box $P_1\times\ldots\times P_k$ has a time-ordered local model with respect to any bipartition $\{A|B\}$ of $\{1,\ldots,kn\}$. We assume without loss of generality that neither A nor B hold any box entirely; as the single boxes are not correlated among each other, boxes held entirely by one party cannot be used to create correlations with the other party. Then, as every P_j has a time-ordered local model for every bipartition of $\{(j-1)n+1,\ldots,jn\}$, A and B can clearly be written as $A=\bigcup_j A_j$ and $B=\bigcup_j B_j$, where $\{A_j|B_j\}$ is an appropriate partition of P_j. For instance, for $n=3$, $k=2$ and $A=\{1,2,4\}, B=\{3,5,6\}$ one would choose $\{A_1|B_1\}=\{\{1,2\},\{3\}\}$ and $\{A_2|B_2\}=\{\{4\},\{5,6\}\}$. Theorem 3.4 then implies the assertion. □

Let us now turn to the main result of this chapter which states that bipartite information principles are not sufficient to single out quantum correlations from the set of nonsignalling correlations. We want to show that the set of fully time-ordered bilocal correlations is strictly larger than the set of quantum correlations. To this end we restrict to the case of $n=3$ parties with binary inputs and outputs. The Bell inequality known as “Guess Your Neighbor’s Input”

$$\beta(P)=P(000|000)+P(110|011)+P(011|101)+P(101|110)\leq 1 \qquad (4.12)$$

is not only fulfilled by all classical correlations but also by quantum correlations (Almeida et al., 2010). That is, a violation of this inequality by some nonsignalling distribution implies that the given distribution is outside the set of quantum correlations.

We have now presented all the necessary ingredients to prove our main result. To demonstrate the existence of supra-quantum correlations that are compatible with any bipartite information principle we maximise the expression (4.12) over tripartite nonsignalling distributions that admit a fully time-ordered bilocal model. This optimisation defines a linear program that can be solved efficiently, see Appendix A. Formally we have

$$\begin{aligned} \beta^\star = \text{maximise} \quad & \beta(P) \\ \text{subject to} \quad & P \text{ fully time-ordered bilocal.} \end{aligned} \tag{4.13}$$

The maximization yields a value of $\beta^\star = 7/6$, implying the existence of tripartite supra-quantum correlations admitting a fully time-ordered bilocal model. Thus, we have the following

Theorem 4.2. *No essentially bipartite information principle is sufficient to single out the set of quantum correlations from the set of nonsignalling correlations in the case of three or more parties.*

To summarise, we have shown that for any $n \geq 3$ there are n-partite nonsignalling correlations that fulfil the principles of information causality and non-trivial communication complexity although they do not belong to the set of quantum correlations.

The presented reasoning actually applies to every other principle based on bipartite information concepts. This result provides a helpful insight for the formulation of a future principle aiming at distinguishing between quantum and supra-quantum correlations. In contrast to the no-signalling principle, such a forthcoming principle will need to be an intrinsically multipartite concept. Therefore, future research should be devoted to the development of information concepts of genuinely multipartite character. This would include in particular the question as to whether non-trivial communication complexity and information causality can be reformulated for the case of n parties to yield truly multipartite principles.

Interestingly, the inequality we used to certify the non-quantumness of the found correlations, the GYNI inequality, is an example of a constraint on correlations derived from the recently introduced multipartite principle of *local orthogonality.* This principle can be interpreted as a multipartite generalisation of the no-signalling principle and gives better approximations on the set of quantum correlations than any other principle known to date (Fritz et al., 2012a,b).

5. Nonlocality in sequential correlation scenarios

We have seen that the study of correlation scenarios allows one to distinguish in an operational way classical, i.e. local, correlations from nonlocal ones. This chapter investigates the phenomenon of nonlocality in the more general situation of sequential correlation scenarios, where distant observers perform a sequence of measurements on their physical systems. One might be tempted to think that such a sequence of measurements could effectively be seen as a bigger single measurement, which would reduce the investigation of sequential correlations to the study of standard correlation scenarios and, therefore, not allow any new conclusions. However, we will see that sequential scenarios are in many ways richer that the ones involving only a single measurement per party at each round.

Popescu (1995), and later Gisin (1996), showed that certain quantum states give only rise to local correlations in scenarios with one measurement per party in each round, but lead to nonlocal correlations when sequences of measurements are performed. These results led Teufel et al. (1997) to investigate the relation between entanglement and local models for certain quantum states in scenarios of sequential measurements in more detail.

Motivated by this manifestation of nonlocality, named "hidden" nonlocality by Popescu (1995), we set out to develop a general framework for the study of nonlocal correlations that arise from sequences of measurements. As in Chapter 3 our approach is operational. Thus, rather than investigating the nonlocal properties of certain quantum states, our task is to identify the set of allowed operations in a sequential correlation scenario so that we can define nonlocality as a resource that cannot be created by these operations. This will lead us to a definition of sequential nonlocality that contains as a particular case standard nonlocality for single measurement rounds and also hidden nonlocality.

5.1. Sequential correlation scenarios

When we introduced the general notion of correlation scenarios in Chapter 2 we assumed that for every physical system, prepared by the common source, each

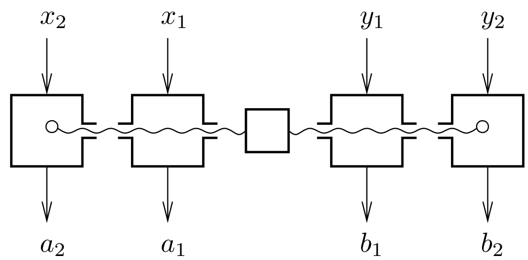


Figure 5.1.: Sequential correlation scenario in the bipartite case. In every run of the experiment a common source prepares a physical system and each of the two parties receives a subsystem. The parties A and B choose the settings x_1 and y_1 for their first measurement respectively and observe the outcomes a_1 and b_1; after recording the outputs of the first measurement the parties choose to perform the measurements x_2 and y_2 yielding outcomes a_2 and b_2. After many runs of the experiment the parties can get together and calculate the correlations, i.e. the joint probabilities $P(a_1a_2b_1b_2|x_1x_2y_1y_2)$ of observing the outcomes a_1, a_2, b_1, b_2 given the measurement settings x_1, x_2, y_1, y_2.

party would choose one measurement to perform and record the corresponding result before the source would generate a new physical system for the next run of the correlation experiment. The data collected in this way allowed the parties when having come together after many runs to calculate the joint probabilities. The current section generalises this idea to the case where the parties perform a sequence of measurements on their part of the system in every run of the experiment.

Let us make more precise what we mean by this type of sequential correlation scenario by first looking at the case of two parties. As in the previously studied situations, a common source produces a bipartite physical system and sends one subsystem to A and the other to B. In contrast to the situations studied before, each party has now not only one set of possible measurement settings but one set of possible settings for each measurement of the sequence of measurements it is going to perform in each run of the experiment. To keep notation simple let us start with the case of a sequence of two measurements for each party as in Fig. 5.1, where we label the i-th measurement setting and the i-th outcome with x_i, a_i and y_i, b_i for A and B. Then, the observed correlations are the collection of joint probabilities

$$P(a_1a_2b_1b_2|x_1x_2y_1y_2). \tag{5.1}$$

Clearly, the outcome of the first measurement cannot depend on later measurement choices; but in the present scenario of sequential measurements later outcomes may well depend on the settings and outcomes of previous measurements. Therefore, the correlations $P(a_1a_2b_1b_2|x_1x_2y_1y_2)$ should not be viewed as four-partite nonsignalling correlations, but rather as a bipartite distribution, where no-signalling holds with respect to A versus B but where signalling from the first measurement of each party to the second of the same party is allowed.

Formally, the no-signalling condition between A and B means that a correlation P that was obtained from a sequence of two measurements for each party obeys

$$\sum_{b_1,b_2} P(a_1a_2b_1b_2|x_1x_2y_1y_2) \qquad \text{independent of } y_1, y_2 \tag{5.2}$$

$$\sum_{a_1,a_2} P(a_1a_2b_1b_2|x_1x_2y_1y_2) \qquad \text{independent of } x_1, x_2, \tag{5.3}$$

which guarantees that the marginal distributions for A and B

$$P_A(a_1a_2|x_1x_2) = \sum_{b_1,b_2} P(a_1a_2b_1b_2|x_1x_2y_1y_2) \tag{5.4}$$

$$P_B(b_1b_2|y_1y_2) = \sum_{a_1,a_2} P(a_1a_2b_1b_2|x_1x_2y_1y_2) \tag{5.5}$$

are well-defined. Furthermore, as later measurements cannot influence the outcome of previous ones, the correlations further have to fulfil

$$\sum_{a_2} P(a_1a_2b_1b_2|x_1x_2y_1y_2) \qquad \text{independent of } x_2 \tag{5.6}$$

$$\sum_{b_2} P(a_1a_2b_1b_2|x_1x_2y_1y_2) \qquad \text{independent of } y_2. \tag{5.7}$$

Correlations fulfilling the above conditions are the objects of interest for the study of sequential correlation scenarios. This notion can straightforwardly be generalised to the case of longer sequences and more than two parties.

Definition 10 (Sequential correlations)**.** For n parties $A_1, \dots, A_n$, where A_i performs a sequence of s_i measurements, let $(a^i_j,\, x^i_j)$ denote the j-th outcome and setting of the i-th party for $1 \leq j \leq s_i$ and $1 \leq i \leq n$. The correlations, given by the collection of the joint probabilities

$$P(\mathbf{a}^1 \dots \mathbf{a}^n|\mathbf{x}^1 \dots \mathbf{x}^n) \tag{5.8}$$

with $\mathbf{a}^i = (a^i_1, \ldots, a^i_{s_i}), \mathbf{x}^i = (x^i_1, \ldots, x^i_{s_i})$, are said to be *n-partite sequential with respect to* $\mathbf{s} = (s_1, \ldots, s_n)$, if for $1 \leq i \leq n$

$$\sum_{a^i_j, \ldots, a^i_{s_i}} P(\mathbf{a}^1 \ldots \mathbf{a}^n | \mathbf{x}^1 \ldots \mathbf{x}^n) \qquad \text{is independent of } (x^i_j, \ldots, x^i_{s_i}) \tag{5.9}$$

for all $1 \leq j \leq s_i$.

Interpreting the pairs $(\mathbf{a}^i, \mathbf{x}^i)$ as the overall output-input pair of party A_i, n-partite sequential correlations as in the above definition constitute an n-partite nonsignalling box. However, the conditional probabilities expressed in the input-output pairs of the individual measurements $\{(a^i_{j_i}, x^i_{j_i}) | 1 \leq j_i \leq s_i\}$ for $1 \leq i \leq n$ need in general not fulfil the no-signalling principle, if seen as the collection of the probabilities of the s variables $\{a^i_{j_i}\}$ conditioned on the s variables $\{x^i_{j_i}\}$, where $s = \sum_i s_i$ the total number of measurements performed in one run of the experiment. The signalling displayed by the distribution when expressed in the individual inputs and outputs is constrained by Eq. (5.9), which allows signalling only among the measurements of the same party in such a way that outcomes can depend at most on all the settings chosen by that party so far.

Ignoring the length of the input and output alphabets for each measurement, the correlation scenario is then characterised by the number of parties n and the vector $\mathbf{s} = (s_1, \ldots, s_n)$ specifying the length of the sequence of measurements for each party.

5.2. Hidden nonlocality

Before we turn to the task of defining the general notion of locality in a sequential correlation scenario, we discuss in this section a known example of how the nonlocality of quantum states can be revealed in a sequential correlation scenario.

The only known way to generate nonlocal correlations consists in measuring entangled quantum states and it is known that every pure bipartite entangled state violates some Bell inequality (Gisin, 1991). However, Werner (1989) showed that this is not the case for mixed states; he constructed a class of bipartite mixed states that are entangled and allow for a local hidden-variable model. These *Werner states* W act on $\mathbb{C}^d \otimes \mathbb{C}^d$ and are of the form

$$W = p\frac{\mathbb{I} + \mathbb{F}}{d(d+1)} + (1-p)\frac{\mathbb{I} - \mathbb{F}}{d(d-1)}, \tag{5.10}$$

where $\mathbb{I}$ denotes the identity matrix on the $d \times d$ dimensional Hilbert space, $\mathbb{F} = \sum_{ij} |i\rangle \langle j| \otimes |j\rangle \langle i|$ the flip operator, and $0 \leq p \leq 1$. For $p = \frac{1+d}{2d^2}$ these states are entangled but do not violate any standard Bell inequality. Werner proved this by explicitly constructing a local hidden-variable model that can reproduce the correlations of these states for arbitrary projective measurements (Werner, 1989).

However, Popescu (1995) noted that a state as in Eq. (5.10) can give rise to correlations that are incompatible with an explanation by local hidden-variables if it is subjected to a sequence of measurements, thereby revealing what he named "hidden" nonlocality. To see how this argument works, suppose that the system is subjected to a sequence of two projective measurements for each party. First, each party performs a measurement that corresponds to the projection of that party's subsystem onto a two-dimensional subspace (or its orthogonal complement), i.e. A performs the measurement given by the projections $\{P, \mathbb{I}_d - P\}$ and B the measurement given by the projections $\{Q, \mathbb{I}_d - Q\}$, where

$$P = |1\rangle\langle 1|_A + |2\rangle\langle 2|_A \tag{5.11}$$

$$Q = |1\rangle\langle 1|_B + |2\rangle\langle 2|_B \,. \tag{5.12}$$

Now, after recording the outcome of the first measurements the parties choose their measurement settings for the second round of measurements. When the parties obtained the outcomes corresponding to the projections P and Q respectively in the first round, the post-measurement state is given by

$$W' = \frac{P \otimes Q W P \otimes Q}{\mathrm{tr}(P \otimes Q W)} \tag{5.13}$$

$$= \frac{d}{d+2} \left(\frac{1}{2d} \mathbb{I}_4 + \left|\Psi^-\right\rangle\left\langle\Psi^-\right| \right), \tag{5.14}$$

where

$$\left|\Psi^-\right\rangle = \frac{1}{\sqrt{2}}(|1\rangle\,|2\rangle - |2\rangle\,|1\rangle) \tag{5.15}$$

the singlet state. If the parties now choose observables A_0, A_1, B_0, B_1 for A and B respectively that give the maximal value of the CHSH expression for the singlet state, they obtain the following value

$$\beta = \mathrm{tr}(W'(A_0B_0 + A_0B_1 + A_1B_0 - A_1B_1)) = \frac{2\sqrt{2}d}{d+2} \tag{5.16}$$

that depends on the local dimension d. For $d \geq 5$ we have $\beta > 2$ indicating that in this case the observed correlations cannot be explained by a local hidden-variable model. So, even though a local hidden-variable model can account for all the correlations obtained from Werner states with $d \geq 5$ that result from a single round of local projective measurements, no such model could account for the correlations obtained when a sequence of two measurements is performed by each party.

The example Popescu (1995) gave was concerned with a specific class of quantum states. Clearly, those states display some sort of nonlocality but they have a standard local hidden-variable model. Thus, the question naturally arises how to formulate locality in sequential correlation scenarios. We are therefore interested in defining nonlocality in such scenarios operationally and constructing local models that do not display Popescu's "hidden" nonlocality.

Another example of hidden nonlocality was reported by Gisin (1996). He found examples of entangled states in dimension $d = 2$ that do not violate the CHSH inequalities for rounds of single measurements but do so when sequences of generalised measurements, given by POVMs, are performed.

For the sake of simplicity let us for now focus on the case considered by Popescu, i.e. a sequence of two measurements for each party, where the first measurement by each party is always the same. We will denote by x_2, y_2 the measurement settings for the second measurement and by a_i, b_i for $i = 1, 2$ the outcome of the i-th measurement of the parties A, B.

Obviously, the notion of locality in the current scenario of sequential measurements should include the standard notion of locality in the sense of Bell, that is a probability distribution fulfilling Eqs. (5.2), (5.3), (5.6) and (5.7) should be able to be decomposed as

$$P(a_1 a_2 b_1 b_2 | x_2 y_2) = \sum_\lambda p_\lambda P_A^\lambda(a_1 a_2 | x_2) P_B^\lambda(b_1 b_2 | y_2). \tag{5.17}$$

After the discussion of Popescu's example it is also clear, that another necessary condition for an appropriate definition of locality in sequential scenarios is the absence of hidden nonlocality. Therefore, one will further require, that all possible post-selections have a local model as well, i.e.

$$P(a_2 b_2 | x_2 y_2 a_1 b_1) = \sum_\lambda p_\lambda^{a_1 b_1} P_A^\lambda(a_2 | x_2) P_B^\lambda(b_2 | y_2) \tag{5.18}$$

for all values of (a_1, b_1) and where the weights $p_\lambda^{a_1 b_1}$ will in general depend on the outputs of the first measurement round.

Let us return to the explicit example by Popescu. Denote the first measurements of A and B by the projectors P_{a_1} and Q_{b_1}; and the second measurements by $\tilde{P}_{a_2|x_2}$ and $\tilde{Q}_{b_2|y_2}$. Then the probabilities read

$$P(a_1 a_2 b_1 b_2 | x_2 y_2) = \mathrm{tr}\left(P_{a_1} \tilde{P}_{a_2|x_2} P_{a_1} \otimes Q_{b_1} \tilde{Q}_{b_2|y_2} Q_{b_1} W\right). \tag{5.19}$$

Now, for Popescu's example the projections of the first and second measurement commute for both A and B. Thus, the expression in Eq. (5.19) can be seen as correlations arising from a single projective measurement on each side and are therefore, due to the explicit hidden-variable model constructed by Werner, local in the sense of Eq. (5.17). On the other hand, they do not fulfil the condition of Eq. (5.18) for the probabilities post-selected on the first outcome of the first measurement,

$$P(a_2b_2|x_2y_2, a_1 = 1, b_1 = 1), \tag{5.20}$$

violate the CHSH inequality.

As said, the first condition (5.17) is nothing but the standard locality condition in the spirit of Bell between the two parties A and B when the pairs (a_1, a_2) and (b_1, b_2) are seen as one output for A and B respectively. The second condition (5.18) ensures that there is no hidden nonlocality as the correlations that arise from the subsequent measurement can be simulated by a local hidden-variable model whatever results were obtained in the first measurement round. As we will see in the following, these necessary requirements are in general not sufficient to capture the notion of locality in a sequential correlation scenario.

5.3. Operational definition of nonlocality for sequential correlations

Popescu's example already showed that the standard notion of locality is not sufficient to capture the behaviour of correlations that can arise in a sequential correlation scenario. Motivated by Popescu's result and our findings from Chapter 3 we set out to give an operational definition of nonlocality when dealing with sequences of measurements. The discussion of nonlocality as a resource in Chapter 3 showed the importance of wirings for a consistent definition of nonlocal correlations in operational terms. Therefore, when trying to give an operational definition in the present scenario, we must ask what kind of wirings are allowed within a sequential correlation scenario and what properties the resulting correlations have.

As already mentioned in the remark following Definition 10, sequential correlations given as the collection of joint probabilities $P(\mathbf{a}^1 \dots \mathbf{a}^n|\mathbf{x}^1 \dots \mathbf{x}^n)$ constitute an n-partite nonsignalling box, when interpreting $(\mathbf{a}^i|\mathbf{x}^i)$ as the output-input pair of party A_i for $1 \leq i \leq n$. However, the no-signalling conditions do in general not hold among all the $s = \sum_i s_i$ variables $(a^i_{j_i}|x^i_{j_i})$ corresponding to all measurement outputs given the inputs for one run of the experiment. Thus,

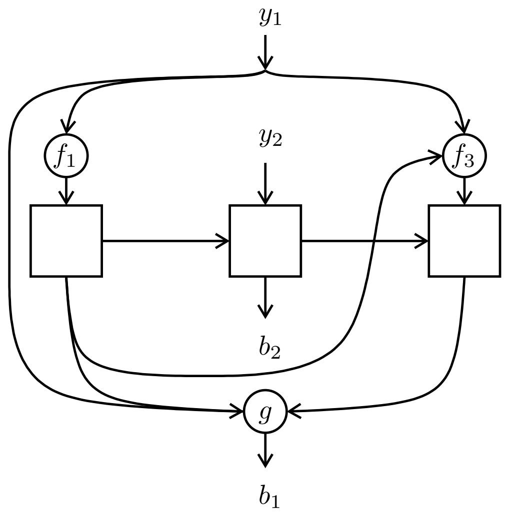


Figure 5.2.: Wiring of sequential correlations. Given the sequential correlations resulting from successive measurements of one party, one can define wired correlations by specifying functions f_1, f_3, g. If the original sequential correlations were described by $P(a_1a_2a_3|x_1x_2x_3)$, then the wired correlations are characterised by the probabilities $\tilde{P}(b_1b_2|y_1y_2) = \sum_{a_1,a_3} P(a_1b_2a_3|f_1(y_1)y_2f_3(a_1,y_1))\delta^{b_1}_{g(y_1,a_1,a_3)}$. However, in the shown case the resulting correlations $\tilde{P}$ are no longer sequential as the wiring makes use of the first and third measurement. To ensure that the resulting correlations are again sequential a valid wiring for sequential correlations must involve strictly successive measurements as in Definition 11.

a valid wiring for a sequential correlation scenario must respect this temporal order of the measurements.

So, given an n-partite sequential scenario we are now interested in the wirings one party A_i can apply among the s_i measurements it performs. Clearly, a wiring can in every step only make use of the inputs and outputs of earlier measurements. But if the party does not use all its measurements for the wirings and we want the resulting correlations to be again sequential correlations, then the causal independence relations of Eq. (5.9) impose further restrictions.

To see this, let us for simplicity consider that party A_i performs a sequence of three measurements, i.e. we consider the probabilities $P_{A_i}(a_1a_2a_3|x_1x_2x_3)$. Now, assume the wiring makes use of the first and the last measurement of the sequence as in Fig. 5.2, i.e. A_i specifies functions f_1, f_3, g to be applied on P_{A_i} to yield the wired correlations

$$\tilde{P}(b_1b_2|y_1y_2) = \sum_{\substack{a_1,a_3 \\ \text{s.t. } g(y_1,a_1,a_3)=b_1}} P_{A_i}(a_1b_2a_3|f_1(y_1)y_2f_3(a_1,y_1)). \tag{5.21}$$

As this wiring respects the temporal order of the measurements, $\tilde{P}$ is a well-behaved probability: if one treats (b_1, b_2) as a single output and (y_1, y_2) as the input, one has $\sum_{b_1,b_2} \tilde{P}(b_1b_2|y_1y_2) = 1$. But if one wants to interpret $\tilde{P}$ as correlations obtained from a sequence of two measurements, one observes that it is not clear that the causal independence relations are met, as

$$\sum_{b_2} \tilde{P}(b_1b_2|y_1y_2) = \sum_{\substack{a_1,a_3,b_2 \\ \text{s.t. } g(y_1,a_1,a_3)=b_1}} P_{A_i}(a_1b_2a_3|f_1(y_1)y_2f_3(a_1,y_1)) \tag{5.22}$$

will in general depend on y_2. Therefore, these correlations can no longer be interpreted as the result of a sequential correlation experiment. Thus, a valid wiring in a sequential correlation scenario must be built from wirings that use only strictly successive measurements in the temporal order specified by the scenario. This leads us to the following

Definition 11 (Sequential wiring)**.** Let P be n-partite sequential correlations with respect to $\mathbf{s} = (s_1, \ldots, s_n)$. A *sequential wiring* for party A_i is specified by $l \in \{1, \ldots, s_i\}$, $p \in \{0, \ldots, s_i - l + 1\}$ and functions $f_1, \ldots, f_p, g$ and takes P to the correlations P'. If one sets $\mathbf{t} = (s_1, \ldots, s_{i-1}, s_i - p + 1, s_{i+1}, \ldots, s_n)$ and denotes the outputs and inputs for the wired correlations P' by

$$\mathbf{b}^j = (b^j_1, \ldots, b^j_{t_j}) \qquad \mathbf{y}^j = (y^j_1, \ldots, y^j_{t_j}) \tag{5.23}$$

for $1 \leq j \leq n$, then P' is characterised by the probabilities

$$\begin{aligned} &P'(\mathbf{b}^1 \ldots \mathbf{b}^n|\mathbf{y}^1 \ldots \mathbf{y}^n) \\ = &\sum_{\substack{a_1,\ldots,a_p \\ \text{s.t. } g(y^i_l,a_1,\ldots,a_p)=b^i_l}} P(\mathbf{b}^1 \ldots \mathbf{b}^{i-1}\mathbf{b}'\mathbf{b}^{i+1} \ldots \mathbf{b}^n|\mathbf{y}^1 \ldots \mathbf{y}^{i-1}\mathbf{y}'\mathbf{y}^{i+1} \ldots \mathbf{y}^n), \end{aligned} \tag{5.24}$$

where

$$\mathbf{b}' = (b^i_1, \ldots, b^i_{l-1}, a_1, \ldots, a_p, b^i_{l+1}, \ldots, b^i_{t_i}) \tag{5.25}$$

$$\mathbf{y}' = (y^i_1, \ldots, y^i_{l-1}, f_1(y^i_l), f_2(y^i_l, a_1), \ldots, f_p(y^i_l, a_1, \ldots, a_{p-1}), y^i_{l+1}, \ldots, y^i_{t_i}). \tag{5.26}$$

We argued that this definition of a sequential wiring would again result in n-partite sequential correlations. That this is indeed the case is easily seen by the following

Proposition 5.1. *The correlation P' from Definition 11 are sequential with respect to* $\mathbf{t}$.

Proof. As the wiring only affects party A_i, we only have to check the causal independence conditions for the marginal distribution

$$P'_{A_i}(\mathbf{b}^i|\mathbf{y}^i) = \sum_{\{\mathbf{b}^j|j\neq i\}} P'(\mathbf{b}^1 \dots \mathbf{b}^n|\mathbf{y}^1 \dots \mathbf{y}^n) \tag{5.27}$$

$$= \sum_{\substack{a_1,\dots,a_p \\ \text{s.t. } g(y_l^i,a_1,\dots,a_p)=b_l^i}} P_{A_i}(\mathbf{b}'|\mathbf{y}') \tag{5.28}$$

with $\mathbf{b}'$ and $\mathbf{y}'$ as in the definition. We need to show that

$$\sum_{b_m^i,\dots,b_{t_i}^i} P'_{A_i}(\mathbf{b}^i|\mathbf{y}^i) \qquad \text{independent of } (y_m^i,\dots,y_{t_i}^i) \tag{5.29}$$

for $1 \leq m \leq t_i$. If $m > l$, then (5.29) obviously holds. Now, if $m \leq l$, then the condition in terms of the function g for the sum (5.28) is automatically fulfilled and the fact that P_{A_i} is sequential implies the desired independence relations. □

Thus, as a sequential wiring results in correlations that are again sequential, the above procedure can be successively applied, for the same or any other party, to yield further wirings.

Motivated by Popescu's result and the discussion of nonlocality as a resource in Chapter 3 we also consider the operation of post-selections on sequential correlations. Given the scenario, it is clear that one party can only post-select on series of strictly successive measurements starting with its first one.

Definition 12 (Sequential post-selection)**.** Let P be n-partite sequential correlations with respect to $\mathbf{s} = (s_1,\dots,s_n)$. Conditioning P by the j-th party on its first p outcomes to be $(a'_1,\dots,a'_p) = \mathbf{a}'$ given the settings $(x'_1,\dots,x'_p) = \mathbf{x}'$ defines the *post-selected correlations* P', where P' are n-partite sequential correlations with respect to $\mathbf{t} = (s_1,\dots,s_{j-1},s_j - p,s_{j+1},\dots,s_n)$ and the joint probabilities are given by

$$
\begin{aligned}
&P'(\mathbf{a}^1 \dots \mathbf{a}^n | \mathbf{x}^1 \dots \mathbf{x}^n) \\
&\quad = \frac{1}{P_j(\mathbf{a}'|\mathbf{x}')} P(\mathbf{a}^1 \dots \mathbf{a}^{j-1}(\mathbf{a}', \mathbf{a}^j)\mathbf{a}^{j+1} \dots \mathbf{a}^n | \mathbf{x}^1 \dots \mathbf{x}^{j-1}(\mathbf{x}', \mathbf{x}^j)\mathbf{x}^{j+1} \dots \mathbf{x}^n).
\end{aligned}
\tag{5.30}
$$

With post-selection and wirings defined for sequential scenarios, we can now define nonlocality for sequential correlations in operational terms.

Definition 13 (Nonlocality of sequential correlations)**.** Let P be n-partite correlations that are sequential with respect to $\mathbf{s}$. The correlations P are called *sequentially local with respect to* $\mathbf{s}$, if every protocol applying sequential wirings and sequential post-selection results in correlations that are local in the standard sense. Otherwise they are called *sequentially nonlocal*.

5.4. Local-causal models for sequential correlation scenarios

We are now in a similar situation as in Chapter 3. Having defined nonlocality in operational terms for the given scenario, we want to construct local models that are compatible with this operational definition. Let us recall the way the standard notion of local hidden-variable models was derived. As explained in Section 2.3, the theorem of Bell assumes a certain causal structure between the hidden variable λ and the events of measurements x, y and outcomes a, b of two separated parties to derive linear constraints, in the form of inequalities, on the joint probabilities $P(ab|xy)$.

Formally, a *causal structure* is a set of variables V and a set of ordered pairs of distinct variables (x, a) determining that x is a direct cause of a relative to V (Pearl, 2009; Spirtes et al., 2001). A convenient way to represent causal structures is through *directed acyclic graphs* (DAGs), where every variable $x \in V$ is a vertex and every ordered pair (x, a) is represented by a directed edge from x to a.

In the standard Bell scenario of two parties there are the observed variables x, y, a, b and further the hidden variable λ, a common cause of both outputs a and b. Thus, we arrive in this case at the causal structure presented in Fig. 5.3.

Another causal structure we have already encountered in Section 3.3 is the one assumed by Svetlichny for his notion of bilocality, see Fig. 5.4. In this case there are the observed variables x, y, z, a, b, c corresponding to the inputs and outputs of the three parties and further the hidden variable λ, a common

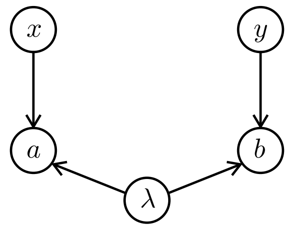


Figure 5.3.: Causal structure of the standard bipartite Bell scenario. The observed variables are the inputs x, y of the two parties and their respective outputs a and b; further a hidden variable λ is assumed as a common cause for both a and b.

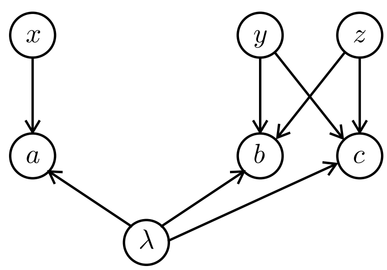


Figure 5.4.: Causal structure underlying Svetlichny's definition of bilocality for tripartite nonsignalling correlations. The observed variables are the inputs x, y, z of the three parties and their respective outputs a, b, c. Furthermore, the hidden variable λ is a common cause of all three outputs and causal influences from y to c and from z to b are allowed. A consistent model defined on this structure has to have a distribution p_λ such that averaging over λ results in tripartite nonsignalling correlations.

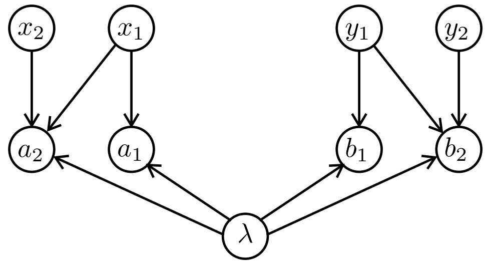


Figure 5.5.: Causal structure for the bipartite sequential correlation scenario with sequences of two measurements for each party. The observed variables are the inputs x_1, x_2 of the first party, the inputs y_1, y_2 of the second party and the corresponding outputs a_1, a_2 and b_1, b_2. The first output of one party is determined by the first input of that party and the hidden variable λ; the second output depends on both inputs of the respective party and the hidden variable λ.

cause for all three outputs. As discussed in Section 3.3, causal influence from z to b and from y to c is allowed in Svetlichny's models, given that the hidden variable is distributed in such a way that after averaging over λ the joint probabilities $P(abc|xyz)$ are nonsignalling. However, we also showed that models that require this fine-tuning of the hidden variable to compensate the signalling present on the level of λ lead to inconsistencies with the operational definition of nonlocality. These inconsistencies led us to define time-ordered local models in Section 3.4 and we could show that these models do not suffer from the above mentioned inconsistencies.

We will now follow a rationale similar to the one that led us to the definition of time-ordered local models when introducing the causal structure for the scenario of sequential measurements. For definiteness let start with the simple case of two parties each performing a sequence of two measurements before turning to the general definition. The observed variables in this case are $x_1, x_2, y_1, y_2, a_1, a_2, b_1, b_2$, where x_i and a_i denote the i-th measurement setting and i-th outcome for A, and y_i and b_i denote the i-th measurement setting and i-th outcome for B; further, as we are interested in the formulation of local models, we assume a hidden variable λ that is a common cause for all outputs.

Clearly, there are direct causal influences from x_i to a_i, from y_i to b_i and from λ to all outputs. As we are treating the parties A and B as separated, we exclude causal influences from inputs of one party to the outputs of the other. Later measurement outcomes of one party, however, will in general depend on earlier settings or outcomes of that party. The response of one party for its i-th measurement should depend only on the given hidden variable λ, the first i measurement settings and the first $i-1$ measurement outcomes of that party. The resulting causal structure is shown in Fig. 5.5. With the corresponding structure for the case of n parties in mind, we can now define local-causal models for sequential correlations.

Definition 14 (Local-causal models)**.** Let P be an n-partite sequential correlation with respect to $\mathbf{s}$ as in Definition 10, described by the joint probabilities $P(\mathbf{a}^1 \dots \mathbf{a}^n | \mathbf{x}^1 \dots \mathbf{x}^n)$. P is said to have a *local-causal model*, if the probabilities can be decomposed as

$$P(\mathbf{a}^1 \dots \mathbf{a}^n | \mathbf{x}^1 \dots \mathbf{x}^n) = \sum_{\lambda} p_\lambda P_1^\lambda(\mathbf{a}^1 | \mathbf{x}^1) \dots P_n^\lambda(\mathbf{a}^n | \mathbf{x}^n), \tag{5.31}$$

where the positive weights p_λ sum to unity and the conditional probabilities P_i^λ are sequential, i.e. for all λ and all i we have

$$\sum_{a^i_j, \dots, a^i_{s_i}} P_i^\lambda(a^i_1 \dots a^i_{s_i} | x^i_1 \dots x^i_{s_i}) \qquad \text{independent of } (x^i_j, \dots, x^i_{s_i}) \tag{5.32}$$

for $1 \leq j \leq s_i$.

Note, any collection of conditional probabilities $\{P_i^\lambda | 1 \leq i \leq n, \lambda \in \Lambda\}$ fulfilling the conditions of Eq. (5.32) defines via Eq. (5.31) valid n-partite sequential correlations admitting a local-causal model for any distribution p_λ of the hidden variable. Once we fix the causal structure, expressed in the conditions Eq. (5.32), no fine-tuning of the the model parameter p_λ is needed to obtain correlations with the correct causal independence conditions. This is in stark contrast to the situation of Svetlichny-bilocal models: there, the distribution p_λ of the hidden variable requires fine-tuning to yield correlations fulfilling the causal independence relations of the given scenario, namely that the resulting correlations be tripartite nonsignalling.

The fact that the models defined on the causal structure shown in Fig. 5.5 or, in the general case of n parties, by the conditions of Eqs. (5.31) and (5.32) do not require fine-tuning makes them the natural choice to study nonlocality in sequential scenarios. Indeed, one can easily see that such models are compatible with the operational definition of sequential locality as defined in Definition 13.

Proposition 5.2. *Let P be n-partite sequential correlations admitting a local-causal model with respect to $\mathbf{s} = (s_1, \dots, s_n)$, then the following holds:*

(i) All correlations obtained from P by post-selection admit a local-causal model.

(ii) Any sequential wiring on P results in correlations that allow for a local-causal model.

Proof. To show (i), one can basically employ the same proof as in Theorem 3.4. We only need to consider post-selection on the first measurement for one party, as the resulting local-causal model will allow to obtain post-selections on more outputs or by other parties by applying post-selection on the first measurement on the resulting correlations successively. So let the first output of the i-th party be $\tilde{a}$ given the setting $\tilde{x}$. The post-selected box P' is characterised by

$$\begin{aligned} &P'(\mathbf{a}^1 \dots \mathbf{a}^n | \mathbf{x}^1 \dots \mathbf{x}^n) \\ &= \frac{1}{P(\tilde{a}|\tilde{x})} P(\mathbf{a}^1 \dots \mathbf{a}^{i-1}(\tilde{a}, \mathbf{a}^i)\mathbf{a}^{i+1} \dots \mathbf{a}^n | \mathbf{x}^1 \dots \mathbf{x}^{i-1}(\tilde{x}, \mathbf{x}^i)\mathbf{x}^{i+1} \dots \mathbf{x}^n), \end{aligned} \tag{5.33}$$

where

$$\begin{aligned} &P(\tilde{a}|\tilde{x}) \\ &\quad = \sum_{\{\mathbf{a}^j\}} P(\mathbf{a}^1 \dots \mathbf{a}^{i-1}(\tilde{a}, \mathbf{a}^i)\mathbf{a}^{i+1} \dots \mathbf{a}^n | \mathbf{x}^1 \dots \mathbf{x}^{i-1}(\tilde{x}, \mathbf{x}^i)\mathbf{x}^{i+1} \dots \mathbf{x}^n) \end{aligned} \tag{5.34}$$

is well defined due to Eq. (5.32). As P is assumed to have a local-causal model

$$P = \sum_\lambda p_\lambda P_1^\lambda \dots P_n^\lambda \tag{5.35}$$

the marginal for a given λ of the first output $P_i^\lambda(\tilde{a}|\tilde{x})$ is well-defined as well, and we can define new weights

$$p'_\lambda = p_\lambda \frac{P_i^\lambda(\tilde{a}|\tilde{x})}{P(\tilde{a}|\tilde{x})} \tag{5.36}$$

to get a local-causal model for P'

$$P' = \sum_\lambda p'_\lambda P_1^\lambda \dots P_{i-1}^\lambda P'^\lambda P_{i+1}^\lambda \dots P_n^\lambda, \tag{5.37}$$

with respect to $\mathbf{t} = (s_1, \dots, s_{i-1}, s_i - 1, s_{i+1}, \dots, s_n)$ where the conditional probability P'^λ is the post-selection of P_i^λ on $(\tilde{a}|\tilde{x})$, i.e.

$$P'^\lambda(\mathbf{a}^i|\mathbf{x}^i) = \frac{1}{P_i^\lambda(\tilde{a}|\tilde{x})} P_i^\lambda(\tilde{a}\mathbf{a}^i|\tilde{x}\mathbf{x}^i). \tag{5.38}$$

To see (ii), just observe that in a local-causal model the correlations P_i^λ appearing in the decomposition are by definition sequential, i.e. there is no fine-tuning on the level of λ. Thus, by Proposition 5.1 the wiring can be applied for every λ on the P_i^λ to yield the local-causal model for the wired correlations. □

The result of Popescu formulated in terms of local-causal models as presented here then reads: Werner states of dimension $d \geq 5$ give rise to sequential correlations that do not admit a local-causal model with respect to $\mathsf{s} = (2, 2)$.

5.5. Detection of sequential nonlocality

With the local-causal models for sequential scenarios defined, the question naturally arises whether they allow to detect nonlocality of correlations in more cases than would be possible using standard Bell tests and Popescu's idea of hidden nonlocality.

In other words, are there correlations that are local in the standard notion, do not display hidden nonlocality, but nevertheless cannot be explained by a model as defined in Definition 14?

To give a first answer to this question we will consider the simplest non-trivial case of sequential measurements in a bipartite scenario, namely one measurement for party A and a sequence of two measurements for B, where for each measurement the respective party can choose from two settings yielding one of two possible outcomes. We denote the outcomes of A and B by $a, b_1, b_2 \in \{0, 1\}$ and the inputs by $x, y_1, y_2 \in \{0, 1\}$ and consider the joint probabilities $P(ab_1b_2|xy_1y_2)$.

Let SeqLoc denote the set of sequential correlations P that admit a local-causal model for the given scenario, i.e. for $P \in \mathsf{SeqLoc}$ we have

$$P(ab_1b_2|xy_1y_2) = \sum_\lambda p_\lambda P_A^\lambda(a|x) P_B^\lambda(b_1b_2|y_1y_2) \tag{5.39}$$

with

$$\sum_{b_2} P_B^\lambda(b_1b_2|y_1y_2) \qquad \text{independent of } y_2. \tag{5.40}$$

Further, let $\mathsf{PostLoc}$ denote the set of sequential correlations P that have a local hidden-variable model with respect to $A|B$ and whose post-selected correlations are local as well, i.e. for $P \in \mathsf{PostLoc}$ we have

$$P(ab_1b_2|xy_1y_2) = \sum_\lambda p_\lambda P_A^\lambda(a|x) P_B^\lambda(b_1b_2|y_1y_2) \tag{5.41}$$

and

$$P(ab_2|xy_2b_1y_1) = \sum_\lambda p_\lambda^{b_1|y_1} P_A^\lambda(a|x) P_B^\lambda(b_2|y_2). \tag{5.42}$$

Both SeqLoc and $\mathsf{PostLoc}$ are convex polytopes, that is compact convex sets with a finite number of extreme points. By Proposition 5.2 correlations from SeqLoc do not display hidden nonlocality, so that we have the inclusion $\mathsf{SeqLoc} \subseteq \mathsf{PostLoc}$. Next we will show that this inclusion is in fact strict, i.e. there are correlations P that are in $\mathsf{PostLoc}$ but not in SeqLoc.

In general a convex polytope can be either described by its extreme points or equivalently by the set of facet-defining half-spaces. These half-spaces are given by linear inequalities

$$\beta(P) = \sum_{a,b,c,x,y,z} \beta_{abc|xyz} P(abc|xyz) \leq 1. \tag{5.43}$$

Using standard polytope software we fully characterized the polytope SeqLoc in terms of its facet-defining inequalities, see Appendix B for details. The problem of deciding whether $\mathsf{SeqLoc} \subsetneq \mathsf{PostLoc}$ or $\mathsf{SeqLoc} = \mathsf{PostLoc}$ can then be cast into a set of linear programs maximising the inequalities of SeqLoc over probability distributions from $\mathsf{PostLoc}$.

As it turns out, the following facet defining inequality of the polytope SeqLoc

$$\begin{aligned}\beta(P) = &-P(101|000) + P(101|100) - P(110|001)\\ &+P(001|111) + P(011|110) + P(100|011)\\ &+P(110|011) + P(110|101) - P(111|010)\\ &+P(111|101) \leq 1\end{aligned} \tag{5.44}$$

can be violated by probability distributions from $\mathsf{PostLoc}$:

$$\beta_{\mathsf{PostLoc}} = \max_{\mathsf{PostLoc}} \beta(P) = \frac{3}{2}, \tag{5.45}$$

showing that $\mathsf{SeqLoc} \subsetneq \mathsf{PostLoc}$. The correlations $P^* \in \mathsf{PostLoc}$ attaining the maximum in Eq. (5.45) have by definition a standard local decomposition with respect to $A|B$ and do not display hidden nonlocality. However, the violation of Eq. (5.44) by P^* demonstrates that these correlations cannot be explained by a local-causal model for sequential correlations.

Now, as correlations from SeqLoc are known to be compatible with our operational definition of sequential locality, the violation Eq. (5.45) raises the question whether there is a sequential wiring that takes P^* to bipartite correlations P' that are nonlocal in the standard sense. In fact, we can prove an even stronger result.

Theorem 5.3. *Let P be bipartite sequential correlations with respect to the sequence $\mathbf{s} = (1,2)$, where each measurement has binary inputs and outputs. Then P is sequentially local in the operational sense, if and only if $P \in \mathsf{SeqLoc}$.*

Proof. That P is compatible with the operational definition, if it has a local-causal model, was shown in Proposition 5.2. To see the converse, consider P to be compatible with the operational definition. We have that all post-selections have a local model

$$P(ab_2|xy_2b_1y_1) = \sum_\lambda p_\lambda^{b_1|y_1} P_A^\lambda(a|x) P_B^\lambda(b_2|y_2) \tag{5.46}$$

for $a, b_1, b_2 \in \{-1, 1\}$ and $x, y_1, y_2 \in \{0, 1\}$. Further, for all sequential wirings, specified by functions f_1, f_2, g, the wired correlations

$$P'(ac|xz) = \sum_{\substack{b_1, b_2 \\ \text{s.t. } g(y_1, b_1, b_2) = c}} P(ab_1b_2|xf_1(z)f_2(z, b_1)) \tag{5.47}$$

are local as well. The conditions Eqs. (5.46) and (5.47) are linear constraints on the probabilities of P, so that we can define linear programs

$$\begin{aligned} \beta^\star = \text{maximise} \quad & \beta(P) \\ \text{subject to} \quad & P \text{ fulfils Eqs. (5.46) and (5.47),} \end{aligned} \tag{5.48}$$

for all facet defining inequalities β of SeqLoc. In the present case of just one measurement for A and two for B, these conditions are still tractable and the linear programs can be solved using standard software. We find

$$\beta^\star = \max_{\mathsf{SeqLoc}} \beta(P) \tag{5.49}$$

for all facet defining inequalities β of SeqLoc, which shows that the set of correlations compatible with the operational definition of sequential nonlocality is equal to SeqLoc. □

So, for this simple scenario, where A performs a single measurement and B a sequence of two with binary inputs and outputs for all of them, the local-causal models exactly capture the operational definition of locality. Correlations admitting a local-causal model are not only compatible with the allowed sequential operations, but having such a model is equivalent to be sequentially local in the operational sense.

This result together with the fact $\mathsf{SeqLoc} \subsetneq \mathsf{PostLoc}$, shown above, implies that apart from Popescu's hidden nonlocality there is a new form of nonlocality that can be revealed by studying correlations arising in scenarios of measurement sequences. Formally stated we have the

Theorem 5.4. *In the bipartite sequential scenario with respect to* $\mathbf{s} = (1, 2)$ *with binary inputs and outputs there exist correlations* $P \in \mathsf{PostLoc}$ *that are nonlocal in the operational sense.*

This is certainly a surprising result. There are correlations P that, when seen as bipartite nonsignalling correlations, have a standard local model as in Eq. (5.41) with respect to $A|B$ and do not display hidden nonlocality, however, they allow for the creation of nonlocal bipartite correlations by a local wiring on B's side.

Furthermore, the above result shows that all the different manifestations of nonlocality, be it the standard one, hidden nonlocality, or this new form of sequential nonlocality, can be detected by using the facet inequalities of SeqLoc. A violation of any of these inequalities certifies the presence of one or several forms of nonlocality. This simplifies the analysis considerably. Instead of having to check, for a given sequential correlation P, all post-selections and all protocols involving post-selection and sequential wirings, one only needs to check whether P satisfies all facet inequalities of SeqLoc to decide if P is local.

To summarise, we have introduced a framework for the study of nonlocality where the different parties can perform sequences of measurements in each round of the experiment. Given the importance of nonlocality as resource for device-independent information processing we defined nonlocality in operational terms. Furthermore, we showed that the resulting notion of nonlocality not only contains the known notions of standard nonlocality and Popescu's hidden nonlocality as particular cases but also a new form of sequential nonlocality. In the spirit of Bell's local hidden-variable models we defined local-causal models for these generalised scenarios that capture the operational definition of locality, which allowed us to give a full characterisation of the set of local correlations for a simple sequential scenario. The same analysis can in principle be applied to more complicated scenarios. However, for larger numbers of parties, longer sequences or larger input and output alphabets the problem becomes increasingly intractable. To conclude this chapter let us discuss some of the most important open questions related to nonlocality in sequential correlation scenarios.

One of the most interesting open question with respect to sequential nonlocality concerns the relationship between the set SeqLoc and the set of correlations that are sequentially local in the operational sense. We know that having a local-causal model implies being local in the operational sense, the converse, however, remains an open problem in the case of more general scenarios.

Problem 1. *Let* P *be* n*-partite correlations that are sequential with respect to* $\mathbf{s} = (s_1, \ldots, s_n)$*. Does* P *being sequentially local in the operational sense imply* *that P admits a local-causal model with respect to* $\mathbf{s}$*?*

Suppose a positive answer. Then, for any sequential correlation scenario, the complicated set of operationally local correlations can be characterised by the facet inequalities corresponding to the set SeqLoc and all types of nonlocality for this scenario can be detected by these inequalities. If, however, the answer is negative, then, for some scenario, there are sequential correlations that remain local under all protocols involving wirings and post-selection while lacking a local-causal model.

Another relevant open problem is related to the nonlocality displayed by quantum states. Does this new form of nonlocality open the possibility to detect more quantum states as nonlocal than would be possible with standard Bell tests or using Popescu's argument of hidden nonlocality? Due to the result of Popescu (1995) we know that there are quantum states with a local hidden-variable model for all projective measurements that display hidden nonlocality. But are there quantum states that do not display hidden nonlocality in any sequential scenario but nevertheless give rise to correlations that do not have a local-causal model? If so, this would correspond to a new form of nonlocality exhibited by quantum states going beyond both standard and hidden nonlocality.

Note, however, that there is some ambiguity in the above question. To see this, consider a scenario, where one party performs a sequence of three or more measurements. Now, what exactly is meant when we say that the post-selection on the first outcome of that party should be local? Should it (i) just be local in the standard sense, (ii) in the operational sense with respect to the two or more remaining measurements of that party, or (iii) have a local-causal model for the remaining measurements? In the light of the failure of standard locality for sequential scenarios option (i) should be discarded; whether there is a difference between (ii) and (iii) depends on the answer to Problem 1. Choosing option (iii) one arrives at the following question that was also raised by Teufel et al. (1997). For simplicity we formulate the problem in the bipartite case.

Problem 2. *Let $\mathfrak{H}_1$ and $\mathfrak{H}_2$ be Hilbert spaces of arbitrary dimensions. Is there a quantum state ϱ acting on $\mathfrak{H}_1 \otimes \mathfrak{H}_2$ such that all correlations that can be obtained from ϱ by projective measurements on $\mathfrak{H}_1$ and $\mathfrak{H}_2$ respectively fulfil the following?*

(i) The correlations P obtained for rounds of single measurements have a standard local model, i.e

$$P(ab|xy) = \sum_{\lambda} p_{\lambda} P^{\lambda}(a|x) P^{\lambda}(b|y). \tag{5.50}$$

(ii) The sequential correlations P do not display hidden nonlocality, i.e. the post-selected correlations

$$P'(\mathbf{a}'\mathbf{b}'|\mathbf{x}'\mathbf{y}') = \frac{1}{P(a_1 b_1|x_1 y_1)} P((a_1, \mathbf{a}')(b_1, \mathbf{b}')|(x_1, \mathbf{x}')(y_1, \mathbf{y}')) \quad (5.51)$$

admit a local-causal model.

(iii) The full sequential correlations P do not admit a local-causal model.

Note that this problem is connected to the open question whether generalised measurements in form of POVMs offer an advantage over projective measurements for detecting standard nonlocality of quantum states. A negative answer to this last question together with a positive answer to Problem 1 implies a negative answer to Problem 2. This can be seen as follows.

Proposition 5.5. *Assume a positive answer to Problem 1 and further that every quantum state ϱ that has a standard local model for projective measurements also has such a model for measurements given by POVMs. Then the answer to Problem 2 is negative.*

Proof. We want to show that under the given assumptions the conditions (i),(ii), and (iii) of Problem 2 cannot be all satisfied. We assume (ii) and (iii) and show a contradiction with (i). Assuming (iii) together with the positive answer to Problem 1 implies that there are sequential correlations obtained from ϱ that are sequentially nonlocal in the operational sense. Assuming (ii) only leaves the possibility that there are sequential wirings taking the correlations P to some bipartite nonlocal correlations. Applying such wirings on the sequential correlations obtained from projective measurements, however, defines effective POVMs for both parties. Now, the assumption that POVMs do not offer any advantage over projections implies a contradiction with (i). □

Popescu (1995) already mentioned that his argument using projective measurements to reveal hidden nonlocality does not apply to the case of local dimension $d = 2$. The states found by Gisin (1996) in dimension $d = 2$ do display hidden nonlocality when sequences of generalised measurements in form of POVMs are applied, however, these states do not have a standard local model for all measurements, but are only local in the sense that they do not violate the CHSH inequality for rounds of single measurements.

Teufel et al. (1997) further presented states in dimension $d \geq 3$ that fulfil conditions (i) and (ii) of Problem 2, but they were not able to conclude whether (iii) holds. Based on these findings and the conjecture that entanglement of a quantum state is equivalent to not having a local-causal model they also proposed a scheme for the classification of nonlocality. According to this scheme the nonlocality of a quantum state is characterised by two natural numbers $\langle N, n \rangle$, the *indices of nonlocality.* The first index N denotes the length of the sequences of measurements necessary to reveal the nonlocality, i.e. the smallest number such that the quantum state gives rise to correlations that do not have a local-causal model with respect to $\mathbf{s} = (N, N)$. For instance, pure entangled states have $N = 1$ and Werner states in dimension $d \geq 5$ have $N = 2$; separable states have $N = \infty$. If $N < \infty$, then the second index n denotes the smallest value of N that can be attained by non-trivial measurements. For the case of Werner states in dimension $d \geq 5$ we have $n = 1$, as the post-measurement state has $N = 1$. For states with $N = \infty$ the second index is defined as the minimal number of copies of the state needed to reveal its nonlocality.

Concerning the nonlocality within quantum theory the ultimate goal would be to identify correlation scenarios in which any entangled state displays some sort of non-classical behaviour. With sequential correlation scenarios we presented one possibility to study nonlocality in situations more general than the standard Bell scenario and we further showed that a new form of nonlocality can arise. A different generalisation of the standard scenario was investigated by Fritz (2012). Building on work from Branciard et al. (2012) he considers scenarios with several sources producing physical systems, but no measurement choice for the observers. Whereas every standard Bell scenario can be mapped into such a generalised correlation scenario, most of these correlation scenarios do not correspond to a standard Bell scenario. Fritz (2012) describes examples of nonlocality in several of these generalised scenarios, while raising many open questions at the same time.

6. Nonlocal multipartite correlations from local marginal distributions

In contrast to classical systems, multipartite quantum systems can be entangled and exhibit nonlocal correlations. Beyond their fundamental interest, both properties are resources for quantum information theory (Horodecki et al., 2009; Barrett et al., 2005). It is thus a relevant question to understand the types of quantum states and correlations that are possible in composite quantum systems. This chapter investigates to what extent one can certify the presence of an information resource of a physical system given only partial knowledge of the global state of the system.

Clearly, if a global system does not contain any kind of information resource, nor do its sub-parts. For the case of entanglement as an information resource, it is known that the converse of this statement is in general not true: some non-entangled reduced states are only compatible with global states that are entangled. We extend this result to the case of nonlocality of correlations and provide local marginal correlations that are only compatible with global genuinely multipartite nonlocal correlations. Quantum nonlocality can thus be deduced from the mere observation of local marginal correlations.

6.1. Compatibility conditions of subsystems

In a multipartite system, every subset of parties constitutes a proper system in itself. The fact that these subsystems describe parts of the same total system requires them to satisfy some compatibility conditions. For instance, a bipartite quantum state ϱ_{AB} is compatible with being the reduced state of a tripartite state ϱ_{ABC} if and only if $\varrho_{AB} = \mathrm{tr}_C(\varrho_{ABC})$. While it is straightforward to check whether some reduced states are compatible with a given global state, the question becomes much more intricate when the global state is unknown and one is interested in knowing whether there exists a quantum state compatible with the given marginals. Finding the conditions for compatibility among reduced quantum states that would guarantee the existence of such a global state is known as the quantum marginal problem (Linden et al., 2002; Higuchi et al., 2003; Klyachko, 2004; Hall, 2007). It is the quantum counterpart of the classical marginal problem, which is concerned with the compatibility of marginal probability distributions.

The quantum marginal problem is trivial in the bipartite case: two reduced states, ϱ_A and ϱ_B, are always compatible with the product bipartite quantum state $\varrho_{AB} = \varrho_A \otimes \varrho_B$. However, the situation becomes more interesting when more than two parties are involved. For instance, it is well known that if two parties share a maximally entangled state, then any tripartite quantum state compatible with it must be such that the third party is uncorrelated to the first two. This phenomenon is known as the monogamy of entanglement (Coffman et al., 2000; Koashi and Winter, 2004) and implies that a maximally entangled state $|\phi^+\rangle_{AB}$ is incompatible with any correlated state ρ_{AC} or ρ_{BC}. A similar property, known as the monogamy of nonlocality, is displayed by nonlocal correlations (Barrett et al., 2005). Parts of a system can thus constrain the set of possible global systems in ways that show up in other parts of the same system.

In this chapter we are interested in the question of what can be inferred about the correlations of a global state given only the knowledge of some of its sub-parts. It is clear that if sub-parts of a system display entanglement or nonlocality, so does the global system. However, is the converse also true? For the case of entanglement it is known that the answer to this question is negative: there are separable states of two qubits that are only compatible with entangled multipartite states (Tóth et al., 2007, 2009). To show this, Tóth et al. (2007, 2009) used spin-squeezing-inequalities to detect entanglement and found entangled multi-qubit states whose reduced two-qubit states are separable. As the entanglement criteria they use only rely on two-body correlations, this demonstrates the existence of non-entangled reduced states that are only compatible with entangled global states.

Here we pose a similar question with regard to nonsignalling correlations described by a joint conditional probability distribution. Our goal, then, is to see whether there are local marginal correlations that are only compatible with multipartite nonlocal correlations. We show that this is indeed the case and that, similarly to what happens with entanglement, nonlocality of multipartite correlations can be certified from marginal correlations that admit a local description. We further provide a quantum state and corresponding measurements that exhibit this type of correlations and also demonstrate that the nonlocality present in the full correlations can be genuinely multipartite. Concerning the question of certifying entanglement from separable marginals, we further provide new examples of separable reduced states that are only compatible with an entangled global state.

Our findings show how the compatibility conditions lead to non-trivial results even when acting on a priori useless marginals: it is possible to witness the presence of useful correlations in the global system from useless reduced states.

6.2. Nonlocality from local marginals

As mentioned before in this thesis, nonlocality of quantum correlations represents a property inequivalent to entanglement and has been identified within the paradigm of device-independent quantum information processing as an alternative resource for quantum information protocols. Again, the corresponding scenario consists of different distant observers, each of which can input a classical setting x_i into his part of the system and obtain an output a_i. The correlations of the inputs and outputs are encapsulated in the joint conditional probability distribution $P(a_1 \dots a_n|x_1 \dots x_n)$ that denotes the probability of obtaining the outputs $a_1, \dots, a_n$ when inputs $x_1, \dots, x_n$ are used.

In what follows, we consider a tripartite scenario where we denote the outcomes of the different parties as a, b, c and the inputs as x, y, z; each party is assumed to be able to choose between two measurement settings, labelled as 0 and 1, each of which can have two outcomes, -1 or 1. That is, for $x, y, z \in \{0, 1\}$ and $a, b, c \in \{-1, 1\}$, we are dealing with the joint conditional probabilities $P(abc|xyz)$. It is useful to consider the following parametrization of the probabilities

$$\begin{aligned} P(abc|xyz) =& \frac{1}{8} \left[1 + a\langle A_x\rangle + b\langle B_y\rangle + c\langle C_z\rangle \right. \\ & + ab\langle A_x B_y\rangle + ac\langle A_x C_z\rangle + bc\langle B_y C_z\rangle \\ & \left. + abc\langle A_x B_y C_z\rangle \right], \end{aligned} \tag{6.1}$$

where $\langle A_x\rangle = \sum_a aP(a|x)$ is the expectation value for the outcome of the first party A given input x, $\langle A_x B_y\rangle = \sum_{a,b} abP(ab|xy)$ is the expectation value for the product of the outcomes of A and B given the inputs x and y, and accordingly for the other expressions.

Given the fact that entanglement can be deduced from the observation of separable reduced states only (Tóth et al., 2007, 2009), it seems natural to ask whether one can infer that some tripartite correlations are nonlocal, only from observation of local bipartite marginals. To answer this question in the affirmative one would need to find three bipartite nonsignalling distributions P_{AB}, P_{AC}, P_{BC} that are local but such that any tripartite nonsignalling distribution P_{ABC} compatible with them is nonlocal. Being compatible in this context means that P_{ABC} must have P_{AB}, P_{AC}, P_{BC} as its marginal distributions, i.e. one must have that

$$\sum_c P_{ABC}(abc|xyz) = P_{AB}(ab|xy) \tag{6.2}$$

$$\sum_b P_{ABC}(abc|xyz) = P_{AC}(ac|xz) \tag{6.3}$$

$$\sum_a P_{ABC}(abc|xyz) = P_{BC}(bc|yz). \tag{6.4}$$

Note, that the left hand sides of the above equations are defined independently of the third input, as P_{ABC} is assumed to be nonsignalling. In the following we provide several instances of distributions satisfying these requirements.

First, let us consider the case where we fix the one-party expectation values as

$$\langle A_x \rangle = \langle B_y \rangle = \langle C_z \rangle = \frac{1}{3}, \qquad x, y, z \in \{0, 1\} \tag{6.5}$$

and the two-party expectation values as

$$\langle A_x B_y \rangle = \langle A_x C_y \rangle = \langle B_x C_y \rangle = \begin{cases} +1 & \text{if } x = y = 0, \\ -\frac{1}{3} & \text{otherwise.} \end{cases} \tag{6.6}$$

Clearly, these assignments define the three bipartite marginals unequivocally. It is easy to see that these bipartite correlations fulfil all possible permutations of the CHSH inequality (Clauser et al., 1969), that reads in correlator form for, say, parties A and B

$$-2 \leq \langle A_0 B_0 \rangle + \langle A_0 B_1 \rangle + \langle A_1 B_0 \rangle - \langle A_1 B_1 \rangle \leq 2. \tag{6.7}$$

As this is the only relevant Bell inequality for two parties with binary inputs and outputs (Fine, 1982), it follows that the assignments of Eqs. (6.5) and (6.6) define three local bipartite correlations.

However, only one tripartite nonsignalling distribution has (6.5) and (6.6) as its marginals. To see this, consider any tripartite nonsignalling distribution P_{ABC} that is compatible with the given marginals. The positivity constraints $P_{ABC}(abc|xyz) \geq 0$ together with the fixed values for the one- and two-party expectation values lead to lower bounds on $\langle A_x B_y C_z \rangle$ and $-\langle A_x B_y C_z \rangle$ that uniquely determine the distribution P_{ABC}. As an example let us consider the case of $x = 0, y = z = 1$. The positivity condition $P(abc|011) \geq 0$ for all a, b, c together with Eqs. (6.5) and (6.6) then gives

$$1 + \frac{1}{3}(a + b + c) - \frac{1}{3}(ab + ac + bc) + abc\langle A_0 B_1 C_1 \rangle \geq 0. \tag{6.8}$$

The choices $a, b, c = 1$ and $a, b, c = -1$ then result in $-1 \leq \langle A_0 B_1 C_1 \rangle \leq -1$. Similar conditions can be obtained for the remaining input combinations that ultimately only allow for the assignment

$$\langle A_x B_y C_z \rangle = \begin{cases} +\frac{1}{3} & \text{if } x+y+z \in \{0,1\}, \\ -1 & \text{otherwise.} \end{cases} \tag{6.9}$$

Equations (6.5), (6.6) and (6.9) define an extremal point of the tripartite nonsignalling polytope, the box number 29 in the classification of Pironio et al. (2011). This point is genuinely nonlocal as it violates a Svetlichny-Bell inequality (Pironio et al., 2011). Thus we found a collection of bipartite conditional probabilities that are local, but only compatible with a unique genuinely tripartite nonlocal distribution.

While this first example answers our original question, it is not entirely satisfactory, as no measurements on a quantum system can achieve all bipartite correlations of Eqs. (6.5) and (6.6) at the same time. Indeed, the only possible extension of these correlations, namely box 29 in Pironio et al. (2011), violates the "Guess-Your-Neighbor-Input" inequality Almeida et al. (2010), which is satisfied by quantum correlations.

Let us hence provide a general characterization of marginals that are only compatible with nonlocal probability distributions. To this end, consider the map Φ that projects a tripartite nonsignalling distribution P_{ABC} with binary inputs and outputs to its three bipartite marginal distributions, i.e.

$$\Phi : \mathbb{R}^{2^6} \ni P_{ABC} \mapsto (P_{AB}, P_{AC}, P_{BC}) \in \mathbb{R}^{2^4} \times \mathbb{R}^{2^4} \times \mathbb{R}^{2^4}. \tag{6.10}$$

Then we can define the set Π of bipartite marginal distributions with binary inputs and outputs, that result from a tripartite local nonsignalling probability distribution as

$$\Pi = \{\Phi(P_{ABC}) | P_{ABC} \text{ local}\}. \tag{6.11}$$

Clearly, the set Π is convex and compact and has a finite number of extreme points. It is therefore a convex polytope and can be described by a finite number of inequalities that only involve the marginal distributions P_{AB}, P_{AC}, P_{BC}. If the bipartite marginal distribution of some tripartite nonsignalling correlations violate any of these inequalities, then they cannot be compatible with a local tripartite distribution. Thus, any extension of these marginal distributions to a tripartite nonsignalling distribution must be a nonlocal tripartite distribution. On the other hand, if some bipartite correlations satisfy all the inequalities that define Π, then they are necessarily compatible with some tripartite local correlations.

By replacing in the definition of Π local with bilocal, one can in a similar manner check whether a collection of bipartite marginal distributions are incompatible with a tripartite bilocal distribution using the polytope

$$\Pi' = \{\Phi(P_{ABC}) | P_{ABC} \text{ time-ordered bilocal}\}. \tag{6.12}$$

Here we use the notion of time-ordered bilocal as introduced in Chapter 3. Since the constraints of the polytope Π' are strictly weaker than those of Π, one has $\Pi \subset \Pi'$. Any inequality satisfied by all points of Π' is thus also a valid inequality for Π. An example of an inequality satisfied by Π' is given by

$$\begin{aligned} &-\langle A_0(1 + B_0 + B_1 + C_0)\rangle \\ &-\langle A_1(1 + B_0 + C_0 + C_1)\rangle \\ &-\langle B_0 + C_0 + B_0C_0 + B_1C_1\rangle \leq 4. \end{aligned} \tag{6.13}$$

Violation of this inequality implies that the correlations compatible with the given marginals must be genuinely tripartite nonlocal. Now, this general characterisation allows us to find local bipartite marginal distributions that are only compatible with nonlocal tripartite distributions, where the correlations can be obtained by local measurements on a quantum state.

We choose the three-qubit state

$$\varrho_{\mathrm{W}}(p) = p\, |W\rangle\langle W| + \frac{1-p}{8}\, \mathbb{I}, \tag{6.14}$$

where $|W\rangle = \frac{1}{\sqrt{3}}(|001\rangle + |010\rangle + |100\rangle)$ and $0 \leq p \leq 1$, and measurement settings for the parties A, B, C

$$\begin{aligned} A_0 &= \cos\alpha\sigma_z + \sin\alpha\sigma_x & A_1 &= \cos\alpha\sigma_z - \sin\alpha\sigma_x \\ B_0 &= -\sigma_z & B_1 &= \cos\beta\sigma_z + \sin\beta\sigma_x \\ C_0 &= -\sigma_z & C_1 &= \cos\beta\sigma_z - \sin\beta\sigma_x. \end{aligned} \tag{6.15}$$

For $p > 0.9548$, $\alpha = 3.6241$ and $\beta = 2.0221$ the inequality Eq. (6.13) can be violated.

On the other hand, the reduced states of two parties of $\varrho_{\mathrm{W}}(p)$ are all equal and have the form

$$\varrho_{\mathrm{red}}(p) = \frac{2p}{3}\, |\psi^+\rangle\langle\psi^+| + \frac{p}{3}\, |00\rangle\langle 00| + \frac{1-p}{4}\, \mathbb{I}, \tag{6.16}$$

where $|\psi^+\rangle = 1/\sqrt{2}(|01\rangle + |10\rangle)$. To see that these reduced states only give rise to local correlations in the present scenario, one can use the criterion found by Horodecki et al. (1995).

Theorem 6.1 (Horodecki et al. (1995)). *A two-qubit state ϱ violates the CHSH inequality for some measurements, if and only if $M(\varrho) > 1$, where*

$$M(\varrho) = \max_{x,y}(\|Tx\|^2 + \|Ty\|^2), \tag{6.17}$$

$T_{mn} = \mathrm{tr}(\sigma_m \otimes \sigma_n \varrho)$*, and x, y are mutually orthogonal unit vectors in $\mathbb{R}^3$.*

For the case $\varrho_{\mathrm{red}}(p)$ one finds $M(\varrho_{\mathrm{red}}(p)) < 1$ for every $0 \leq p \leq 1$. Since the CHSH inequality is the only relevant Bell inequality for the case of two parties with binary inputs and outputs, this implies that in the considered scenario all local measurements on $\varrho_W(p)$ lead to correlations that have local bipartite marginal distributions. Thus, in summary, we have obtained an example of local quantum marginal correlations which are only compatible with genuine tripartite nonlocal correlations. A stronger version of our result would consist in finding a collection of reduced states that have a local hidden-variable model for all measurements but whose correlations are at the same time only compatible with global nonlocal correlations.

6.3. Entanglement from separable marginals

Regarding the problem of entanglement detection from separable marginals, note that the global state of a system is known to be generally determinable from its marginals, if one has the promise that the global state is pure. Indeed, consider the bipartite marginals

$$\varrho_{AB} = \varrho_{AC} = \varrho_{BC} = \frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 11|) \tag{6.18}$$

obtained from some three-qubit state ϱ_{ABC}. If the global state is a pure state $|\Psi\rangle\langle\Psi|$, then it follows from its Schmidt-decomposition that it must be the Greenberger-Horne-Zeilinger (GHZ) state

$$|GHZ\rangle = \frac{1}{\sqrt{2}}(|000\rangle + e^{i\phi}\,|111\rangle). \tag{6.19}$$

While these bipartite marginals are separable, the GHZ state is entangled and, thus, observation of separable marginals in this case are only compatible with an entangled pure state.

On the other hand, a sufficiently large number of reduced states of a pure state is almost always enough to uniquely determine the global state among all (pure or mixed) states; with three qubits, the only exception consists of the states that are equivalent under local unitaries to states of the form $a\,|000\rangle +$

$b\,|111\rangle$ (Linden et al., 2002). So, for all other pure entangled states whose marginals are not entangled, entanglement in the global state can be deduced from the observation of separable marginals.

Now, if the global state is not assumed to be pure, then the above analysis immediately fails. For instance, the reduced states of the GHZ state are also compatible with the three-party mixed state

$$\varrho_{ABC} = \frac{1}{2}(|000\rangle\langle 000| + |111\rangle\langle 111|), \tag{6.20}$$

which is separable. Thus, observation of these marginals without further knowledge on the full state does not guarantee entanglement in the whole system. Actually, this result applies to every graph state: for any such state there is always a separable state that has the same two-body reductions (Gittsovich et al., 2010). So no criterion relying on two-particle correlations can detect graph state entanglement.

However, as mentioned before, it was shown that there are separable two-qubit states that are only compatible (among all states) with an entangled global state (Tóth et al., 2009, 2007). Here, we present further examples of this feature involving the reduced states of three-qubit states.

The starting point for our investigation is again a noisy W state. The reduced states

$$\varrho_{\text{red}}\,(p) = \frac{2p}{3}\,|\psi^+\rangle\langle\psi^+| + \frac{p}{3}\,|00\rangle\langle 00| + \frac{1-p}{4}\,\mathbb{I}, \tag{6.21}$$

are separable for $0 \leq p \leq p_{\text{sep}} = 3/\left(1+2\sqrt{5}\right)$. We are interested to see if there exists a value of p with $p \leq p_{\text{sep}}$ such that every three-qubit state compatible with these reductions must be entangled.

To this end, we will look for the maximal value of p such that every three-qubit state having $\varrho_{\text{red}}\,(p)$ as its reductions is not entangled. For simplicity, let us relax this last constraint, allowing the three-qubit state to have a positive partial transposition (PPT) instead of being separable (Peres, 1996). After this relaxation, the maximal value of p corresponds to the solution $p^\star$ of the following semi-definite program (SDP):

$$\begin{aligned} p^\star = \text{maximise} \quad & p \\ \text{subject to} \quad & \varrho \succeq 0, \\ & \operatorname{tr}_X \varrho = \varrho_{\text{red}}\,(p) \text{ for } X = A, B, C \\ & \varrho^{\top_X} \succeq 0 \text{ for } X = A, B, C, \end{aligned} \tag{6.22}$$

where $\varrho^{\top_X}$ denotes the partial transposition of ϱ with respect to the subsystem X. Note that the normalization condition $\operatorname{tr}(\varrho) = 1$ is ensured by the constraints on the bipartite marginals $\operatorname{tr}_X \varrho$.

For every semi-definite program one can always define the dual problem, which is a minimisation problem, if the primal is a maximisation or vice versa. For the primal Eq. (6.22) the dual reads

$$\begin{aligned} d^\star = \text{minimise} \quad & \frac{1}{4}\,\mathrm{tr}(\mu_A + \mu_B + \mu_C) \\ \text{subject to} \quad & \nu_X \succeq 0, \\ & \mathrm{tr}\left[(\mu_A + \mu_B + \mu_C)M\right] \leq -1, \\ & \sum_X \mathbb{I}_X \otimes \mu_X - \nu_X^{\top_X} \succeq 0. \end{aligned} \tag{6.23}$$

where μ_X are 4×4 matrices and ν_X are 8×8 matrices; the expression $\mathbb{I}_X \otimes \mu_X$ denotes the operator that acts as the identity on particle X and as μ_X on the rest.

From weak duality one always has the relation $d^\star \geq p^\star$; every feasible point of the primal problem gives a lower bound $p' \leq p^\star$ and every feasible point of the dual gives an upper bound $d' \geq d^\star$.

Appendix C provides details of the semi-definite programs. There, we solve the primal and dual problem and find variables ρ, μ_X, ν_X that satisfy all the constraints of Eqs. (6.22) and (6.23), while yielding the same bounds $p' = d' = 3/2(2 + \sqrt{17})$. Thus, we find

$$p^\star = d^\star = \frac{3}{2}(2 + \sqrt{17}) \tag{6.24}$$

as the solution of the optimisation problem. Hence, the reduced states $\varrho_{\mathrm{red}}(p)$ of Eq. (6.16) with $p^\star < p \leq p_{\mathrm{sep}}$ certify the presence of entanglement in the global state despite being separable.

The above considerations can be generalised to the case of more than three parties. Again, starting from the noisy W state of n qubits we found a similar behaviour: one can choose separable two-party states that are only compatible with an entangled global state of n qubits. In the case of n parties we start from the state

$$\varrho(n, p) = p\,|W_n\rangle\langle W_n| + \frac{(1-p)}{2^n}\,\mathbb{I}, \tag{6.25}$$

where $|W_n\rangle = \frac{1}{\sqrt{n}}(|0\ldots01\rangle + |0\ldots010\rangle + \ldots + |10\ldots0\rangle)$. In this case the bipartite reduced states are given by

$$\varrho_{\mathrm{red}}(n, p) = \frac{2p}{n}\left|\psi^+\right\rangle\!\left\langle\psi^+\right| + \frac{p(n-2)}{n}\,|00\rangle\langle 00| + \frac{1-p}{4}\,\mathbb{I} \tag{6.26}$$

and they are separable for $p \leq p_{\mathrm{sep}} = n/(4 - n + 2\sqrt{n^2 - 4n + 8})$.

n	3	4	5	6	7
$p^\star$	0.4899	0.6180	0.7464	0.8279	0.8787
p_{sep}	0.5482	0.7071	0.8050	0.8640	0.9009

Table 6.1.: Values of $p^\star$ and p_{sep} for different number n of parties; the two-body reduced states $\varrho_{\text{sep}}(n,p)$ are separable for $p \leq p_{\text{sep}}$ and $p^\star$ is the solution of the SDP (6.27). For all n there are p with $p^\star < p \leq p_{\text{sep}}$ showing that there always is a collection of separable two-body states only compatible with an entangled n-partite state.

The corresponding semi-definite program reads now

$$\begin{aligned} \text{maximise} \quad & p \\ \text{subject to} \quad & p \geq 0,\ \varrho \succeq 0 \\ & \operatorname{tr}_{i_1,\dots,i_{n-2}} \varrho = \varrho_{\text{red}}(n,p) \\ & \varrho^{\top_S} \succeq 0, \end{aligned} \tag{6.27}$$

for all subsets $\{i_1,\dots,i_{n-2}\}$ of $n-2$ elements of $\{1,\dots,n\}$ and the condition of positive partial transpose is imposed with respect to all possible partitions of the n parties into two groups. As in the case of three parties we find that there is a range of the parameter $p^\star < p \leq p_{\text{sep}}$ such that separable bipartite marginals allow for the certification of entanglement in the global state. Table 6.1 summarises our results for $n \leq 7$. The behaviour of the gap $p_{\text{sep}} - p^\star$ for large n is not known, i.e. it is not clear whether the above result will hold for arbitrary n.

To conclude, we have demonstrated how compatibility constraints among marginal distributions allow one to certify the presence of nonlocal correlations in a global state from marginals that can be explained by a local model. In particular, we have provided examples of local bipartite marginals that are only compatible with nonlocal probability distributions, and even with genuinely tripartite nonlocal distributions. Furthermore, these correlations can be obtained by local measurements on an entangled quantum state. This result reveals that local models reproducing some (local) bipartite marginal correlations can be fundamentally incompatible with each other, since the full correlations representing their joint behaviour admit no such model.

In addition, for the case of entanglement we have presented a collection of three separable two-qubit states that are only compatible with an entangled tripartite state; this result was further generalised to the case of more than three parties. From a general point of view, our work proves how compatibility constraints lead to non-trivial results even when acting on separable or local states.

7. Conclusions

This thesis showed that at a full understanding of nonlocal and quantum correlations requires to consider more general scenarios than the standard bipartite case originally studied by Bell (1964). As nonlocality has been identified as a resource for information processing, our goal was to characterise nonlocal correlations in operational terms.

To define nonlocality as a resource in these more general situations we first studied correlation scenarios of an arbitrary number n of parties where the individual parties can be grouped into any number of subgroups according to some given partition of the n-partite system. We then identified the allowed physical operations for this situation and defined nonlocality as the resource that cannot be created with these operations. While the resulting definition of nonlocality coincides with the standard notion of nonlocality for a partition into n groups, inconsistencies of standard local models with the operational definition arise when considering partitions of the n parties into $k < n$ groups, as we demonstrated for the models of Svetlichny (1987). To overcome these inconsistencies we introduced a new class of local models that are compatible with our operational definition of nonlocality. However, whether the conditions imposed by these models define the largest set of correlations compatible with the operational definition remains an open problem.

The definition of quantum correlations is based on the abstract mathematical formalism of quantum mechanics; in contrast to general nonsignalling correlations or local correlations no physical principle is known that would characterise the set of quantum correlations. In particular, no known principle can explain why certain correlations cannot be realised by quantum means, even though they respect the no-signalling principle. Recently, information-theoretic principles were proposed as a possibility to provide such a general principle that would single out the set of quantum correlations from the larger set of nonsignalling correlations.

By building on results from the operational definition of nonlocality we could show a fundamental limitation of this approach: no principle based on bipartite information concepts is able to characterise the set of quantum correlations. Using a special class of the models we introduced in Chapter 3 we identified tripartite nonsignalling correlations that behave classically in every bipartite sense and, therefore, fulfil any bipartite information principle. As we have shown, these correlations, though, lie outside the quantum set. Thus, truly multipartite principles are required for the characterisation of quantum correlations. The general question as to whether such a general principle exists remains an open question and is the subject of ongoing research.

The only known way to generate nonlocal correlations consists in measuring entangled quantum states. However, when considering correlation scenarios with rounds of single measurements it is known that there are entangled states resulting in local correlations for all measurements, as shown by Werner (1989) for the case of projective measurements and by Barrett (2002) for generalised measurements. Popescu (1995) showed that this gap between nonlocality and entanglement can be narrowed, if one considers sequences of measurements performed by each party. Motivated by this result of hidden nonlocality we introduced a general framework for the study of nonlocality in sequential correlation scenarios. Given its importance as a resource for quantum information processing our goal was to define nonlocality in operational terms. To this end we identified the allowed physical operations for scenarios of sequential measurements and defined nonlocality as the property of correlations that cannot be created by these operations. To characterise the resulting notion of nonlocality we introduced local-causal models in the spirit of Bell's local hidden-variable models (Bell, 1964); we showed that our models are compatible with the operational definition of nonlocality for every sequential correlation scenario. Our approach allowed us to identity a new form of nonlocality, apart from the standard one and hidden nonlocality, that can be revealed in sequential correlation scenarios.

Another difficulty concerning the detection of nonlocality or entanglement arises when one only has partial knowledge of the total correlated system. In particular, we studied the question as to whether the presence of nonlocal correlations can be confirmed by the observation of marginal correlations only, even if these marginals are local. Obviously, this question has a negative answer for the case of a global bipartite system; two given one-party marginals are always compatible with the product distribution which is clearly local. However, such a simple argument does not apply to the multipartite case. We answered the above question for the case of more than two parties in the affirmative by providing a collection of three bipartite local marginal distributions that we proved to be only compatible with a genuine tripartite nonlocal distribution. Furthermore, we showed that correlations with these properties can be obtained by local measurements on an entangled quantum state. Similarly, for the case of entanglement we found collections of bipartite separable states that are only compatible with an entangled global state. This shows that the information resources of nonlocality and entanglement in a large system can be detected by measuring only small subsystems, even if these subsystems do not contain the respective resource.

In summary, by extending the study of nonlocal correlations from the standard bipartite case to multipartite correlation networks this thesis provided several insights for a better understanding of nonlocality, especially with regard to its characterisation in operational terms. As we have seen, these results also have implications for the task of describing quantum correlations by physical principles. Having said that, several questions remain open. For instance, it is in general not known to what extent marginal correlations allow one to determine global properties of the correlated system. Another open question concerns the characterisation of quantum correlations by a general physical principle. As we have seen, no principle based on bipartite concepts is sufficient to single out quantum correlations from the larger set of nonsignalling correlations. Interestingly, the mathematical tool we used to certify that there are fully time-ordered bilocal correlations outside of the quantum set, the GYNI inequality, is an example of an inequality derived from the principle of *local orthogonality*. This recently introduced principle is genuinely multipartite and in several cases gives better bounds on the quantum set than any other principle known to date (Fritz et al., 2012a,b).

To conclude, let us consider the two most important open problems related to the work presented in this thesis. For one thing, there is the problem of characterising the set of correlations that are operationally local. The other question concerns the relation between entanglement and nonlocality of quantum states.

Regarding the first question, we have seen that in correlation scenarios with rounds of single measurements the class of time-ordered local models constitutes a set of correlation compatible with the operational definition of locality. However, even for simple scenarios it is not known whether correlations admitting a time-ordered local model define the largest set compatible with the operational definition. In the case of sequential measurements an analogous question arises. There, we could show for the simplest non-trivial scenario that for correlations being operationally local is equivalent to the existence of a local-causal model. Though, in the general case it remains an open problem whether these two notions are equivalent.

With respect to the relation between entanglement and nonlocality of quantum states, the study of sequential correlation scenarios raises the question under which conditions this new framework allows one to identify quantum states as nonlocal that would seem local when only considering rounds of single measurements or using Popescu's argument of hidden nonlocality. As we have seen, the answer to this question is connected to the problem of whether general measurements offer an advantage over projective measurements for the verification of nonlocal correlations. Ultimately, as quantum entanglement is at the heart of all observed nonlocal correlations, intuitively every entangled state should reveal non-classical correlations in some sense. To formalise this 'in some sense' one needs to go beyond the standard Bell scenario, and probably also beyond the scenarios presented in this thesis. An exciting idea is to investigate scenarios with several sources of hidden variables, as in the work by Fritz (2012), for the case of sequential measurements.

A. Solution of the linear program from Section 4.3

This appendix presents the solution of the linear program we encountered in Section 4.3. This solution also constitutes the tripartite time-ordered local distribution whose existence we claimed in the proof of Proposition 3.5.

Let us turn to the solution of the linear program from Section 4.3. In Eq. (4.13) the problem was given as

$$\begin{aligned} \beta^\star = \text{maximize} \quad & \beta(P) \\ \text{subject to} \quad & P \text{ fully time-ordered bilocal,} \end{aligned} \tag{A.1}$$

where β is the "Guess Your Neighbor's Input" (GYNI) expression (Almeida et al., 2010)

$$\beta(P) = P(000|000) + P(110|011) + P(011|101) + P(101|110). \tag{A.2}$$

P being fully time-ordered bilocal means that P can be decomposed as

$$\begin{aligned} P(a_1a_2a_3|x_1x_2x_3) &= \sum_\lambda p_\lambda^{i|jk} P_i^\lambda(a_i|x_i) P_{j\to k}^\lambda(a_ja_k|x_jx_k) \\ &= \sum_\lambda p_\lambda^{i|jk} P_i^\lambda(a_i|x_i,) P_{j\leftarrow k}^\lambda(a_ja_k|x_jx_k) \end{aligned} \tag{A.3}$$

for $(i,j,k) = (1,2,3),(2,3,1),(3,1,2)$, with the distributions $P_{j\to k}^\lambda$ and $P_{j\leftarrow k}^\lambda$ obeying the conditions

$$\sum_{a_k} P_{j\to k}^\lambda(a_ja_k|x_jx_k) = P_{j\to k}^\lambda(a_j|x_j), \tag{A.4}$$

$$\sum_{a_j} P_{j\leftarrow k}^\lambda(a_ja_k|x_jx_k) = P_{j\leftarrow k}^\lambda(a_k|x_k). \tag{A.5}$$

To solve the linear program of Eq. (A.1) one has to list all deterministic strategies of the type (A.3) fulfilling the conditions (A.4) and (A.5) for all three bipartitions. Then, the maximisation can be efficiently solved by varying the corresponding weights $p_\lambda^{i|jk}$ for $(i,j,k) = (1,2,3),(2,3,1),(3,1,2)$ using standard software (Löfberg, 2004).

Let us now present a tripartite distribution P fulfilling all constraints of the linear program and obtaining the maximal value of $\beta(P) = 7/6$. To simplify notation, let us switch from $(a_1a_2a_3)$ to (abc); and from $(x_1x_2x_3)$, to (xyz). Now, consider the tripartite nonsignalling distribution P characterised by the probabilities $P(abc|xyz)$ shown in Table A.1.

	000	001	010	011	100	101	110	111
000	2/3	0	0	0	0	0	0	1/3
001	1/3	1/3	0	0	0	0	1/6	1/6
010	1/3	0	1/3	0	0	1/6	0	1/6
011	1/6	1/6	1/6	1/6	0	1/6	1/6	0
100	1/3	0	0	1/6	1/3	0	0	1/6
101	1/6	1/6	0	1/6	1/6	1/6	1/6	0
110	1/6	0	1/6	1/6	1/6	1/6	1/6	0
111	0	1/6	1/6	1/6	1/6	1/6	1/6	0

Table A.1.: Probabilities $P(abc|xyz)$ attaining the maximum of $7/6$ for the optimisation problem Eq. (A.1), where the rows correspond to the inputs xyz and the columns to the outputs abc.

The value of the "Guess Your Neighbor's Input" expression for P equals

$$\beta(P) = \frac{2}{3} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{7}{6} > 1, \tag{A.6}$$

and thus the correlations of P cannot be obtained by local measurements on a quantum system. To see that P belongs to the set of correlations with a fully time-ordered bilocal model, first notice that $P(abc|xyz)$ is invariant under permutations of the three parties. It is therefore enough to show that it admits a decomposition of the form Eq. (4.9) for the partition $A|BC$. Along this bipartition, probability distributions appearing in the decomposition (4.9) are such that the outcome a only depends on the measurement choice x for every given λ; let a_x denote this outcome for $x = 0,1$. Conditions (A.4) and (A.5) tell us that for every λ the marginal $P_{B\to C}(b|y,\lambda)$ is independent of z, and the marginal $P_{B\leftarrow C}(c|z,\lambda)$ is independent of y. Thus, for $B \to C$ we have that b depends on y and c depends on both z and y. The possible outcomes will then be denoted b_y, c_{yz}. Similarly, for $B \leftarrow C$, the possible outcomes are b_{yz}, c_z.

λ	p_λ	a_0	a_1	b_0	b_1	c_{00}	c_{01}	c_{10}	c_{11}
1	$1/12$	0	0	0	1	0	1	0	1
2	$1/12$	0	0	0	0	0	1	0	1
3	$1/12$	0	0	0	0	0	0	0	1
4	$1/12$	0	0	0	1	0	0	0	1
5	$1/12$	0	1	0	1	0	0	0	0
6	$1/12$	0	1	0	0	0	1	0	0
7	$1/12$	0	1	0	0	0	0	0	0
8	$1/12$	0	1	0	1	0	1	0	0
9	$1/6$	1	0	1	1	1	1	1	0
10	$1/6$	1	1	1	0	1	0	1	1

Table A.2.: Decomposition of the probabilities from Table A.1 into deterministic probability distributions characterized by outcome assignments for the bipartition $A|BC$ in the case $A|B \to C$. For every λ the outcome a only depends on x, and b only depends on y.

λ	p_λ	a_0	a_1	b_{00}	b_{01}	b_{10}	b_{11}	c_0	c_1
1	$1/12$	0	0	0	0	0	1	0	0
2	$1/12$	0	0	0	0	0	1	0	1
3	$1/12$	0	0	0	0	1	1	0	0
4	$1/12$	0	0	0	0	1	1	0	1
5	$1/12$	0	1	0	0	0	0	0	0
6	$1/12$	0	1	0	0	0	0	0	1
7	$1/12$	0	1	0	0	1	0	0	0
8	$1/12$	0	1	0	0	1	0	0	1
9	$1/6$	1	0	1	1	1	0	1	1
10	$1/6$	1	1	1	1	0	1	1	0

Table A.3.: Decomposition of the probabilities from Table A.1 into deterministic probability distributions characterized by outcome assignments for the bipartition $A|BC$ in the case $A|B \leftarrow C$. For every λ the outcome a only depends on x, and c only depends on z.

Tables A.2 and A.3 contain the output assignments corresponding to deterministic probability distributions together with the weights p_λ for $A|B \to C$ and $A|B \leftarrow C$, respectively. Note that, in agreement with Eq. (4.9), the outcome assignments for A and the weights p_λ are the same for both decompositions.

As both tables indeed reproduce $P(abc|xyz)$, the tripartite box P belongs to the set of correlations admitting a fully time-ordered bilocal model. Furthermore, as this decomposition is obviously time-ordered bilocal with respect to the partition $A|BC$, P also constitutes the promised tripartite time-ordered correlations we used in the proof of Proposition 3.5.

B. Inequalities of SeqLoc

This appendix contains the classification of the facet inequalities of the polytope SeqLoc for the scenario discussed in Section 5.5.

We consider a bipartite scenario where A performs a single measurement and B a sequence of two measurements per round; both inputs and outputs are binary for all measurements. Thus, if $P \in \mathsf{SeqLoc}$ for this scenario, it can be decomposed as

$$P(ab_1b_2|xy_1y_2) = \sum_{\lambda} p_\lambda P_A^\lambda(a|x)P_B^\lambda(b_1b_2|y_1y_2) \tag{B.1}$$

with

$$\sum_{b_2} P_B^\lambda(b_1b_2|y_1y_2) \qquad \text{independent of } y_2. \tag{B.2}$$

To find the facet inequalities of the polytope SeqLoc one generates all the deterministic points $P_A^\lambda(a|x)P_B^\lambda(b_1b_2|y_1y_2)$ that fulfil the condition of Eq. (B.2). Given these extremal points as input one can use standard polytope software to obtain all facet defining inequalities. The inequalities for this scenario can be classified into three groups.

Marginals correlations The first group consists of inequalities that only involve one of the marginal distributions $P(ab_1|xy_1)$ or $P(ab_2|xy_1y_2)$. For the marginal $P(ab_1|xy_1)$ the inequalities are equivalent to the CHSH inequalities (up to relabelling of inputs and outputs or interchanging the parties), i.e. we have

$$P(a = b_1|00) + P(a = b_1|01) + P(a = b_1|10) + P(a \neq b_1|11) \leq 3 \tag{B.3}$$

and all its symmetries. In the case of the marginal distribution $P(a_1b_2|x_1y_1y_2)$ the inequalities are again equivalent to the CHSH inequality, but now B can choose among the four different inputs (y_1, y_2). This corresponds to a *lifted* version of the CHSH inequalitiy (Pironio, 2005), i.e. we have in this case the inequality

$$P(a = b_2|000) + P(a = b_2|011) + P(a = b_2|100) + P(a \neq b_1|111) \leq 3. \tag{B.4}$$

and also all its symmetries.

Post-selected correlations The second group comprises facet inequalities involving the post-selected correlations $P(ab_2|xy_2b_1y_1)$. They are again equivalent to the CHSH inequality, i.e. we have for every pair (b_1, y_1) the inequality

$$\begin{aligned} &P(a = b_2|00, b_1y_1) + P(a = b_2|01, b_1y_1) \\ &\qquad + P(a = b_2|10, b_1y_1) + P(a \neq b_2|11, b_1y_1) \leq 3 \end{aligned} \tag{B.5}$$

plus symmetries.

Full correlations The last groups consists of inequalities that involve the full probability distribution $P(ab_1b_2|xy_1y_2)$. In this case we have, again up to symmetries, the inequality

$$\begin{aligned} &-P(101|000) + P(101|100) - P(110|001) \\ &+P(001|111) + P(011|110) + P(100|011) \\ &+P(110|011) + P(110|101) - P(111|010) \\ &+P(111|101) \leq 1. \end{aligned} \tag{B.6}$$

Note, the inequalities from the first group correspond to the condition of standard locality between the parties A and B, the ones from the second group guarantee that P does not display hidden nonlocality. The last group contains the inequalities that allowed us to detect the new form of sequential nonlocality as discussed in Section 5.5.

As mentioned before, the analysis of the situation of sequential measurements presented above can in principle be generalised to longer sequences. However, when increasing the the length of the sequences or the number of inputs and outputs the problem becomes increasingly intractable. In the following we show how an inequality from the third group can be lifted to the case of sequences of two measurements for both A and B.

To this end it is useful to change notation; we now denote the outputs that take values in $\{-1, 1\}$ by (abc) and write (xyz) for (xy_1y_2). Now, consider the following parametrisation of the joint probability distribution:

$$\begin{aligned} P(abc|xyz) = &\frac{1}{8} [1 + a\langle A_x\rangle + b\langle B_y\rangle + c\langle C_{yz}\rangle \\ &+ ab\langle A_xB_y\rangle + ac\langle A_xC_{yz}\rangle + bc\langle B_yC_{yz}\rangle \\ &+ abc\langle A_xB_yC_{yz}\rangle] , \end{aligned} \tag{B.7}$$

where $\langle A_x\rangle = P(a = 1|x) - P(a = -1|x)$ is the expectation value of the outcome for party A given input x; $\langle A_xC_{yz}\rangle = P(ac = 1|xyz) - P(ac = -1|xyz)$ is the expectation value of the product of the outcome of A and the second outcome of B given the inputs x, y, z, and so on.

If one defines the following linear combinations of correlators

$$B = \frac{1}{2}\left[(1 + B_0)C_{01} - (1 - B_0)C_{00}\right] \tag{B.8}$$

$$B' = \frac{1}{2}\left[(1 - B_1)C_{11} + (1 + B_1)C_{10}\right], \tag{B.9}$$

the inequality (B.6) that can be violated with probability distributions from $\mathsf{PostLoc}$ can be written in the compact form

$$\langle A_0(B - B') - A_1(B + B')\rangle \leq 2. \tag{B.10}$$

This expression formally looks like the CHSH inequality. Now consider the situation where also A performs a second measurement choosing a setting $w \in \{0, 1\}$ and obtaining an outcome $d \in \{-1, 1\}$; the correlators corresponding to the second measurement of A are denoted by D_{xw}. Replacing A_0 and A_1 in (B.10) with combinations of correlators similar to (B.8) and (B.9), i.e.

$$A = \frac{1}{2}\left[(1 + A_0)D_{01} - (1 - A_0)D_{00}\right] \tag{B.11}$$

$$A' = \frac{1}{2}\left[(1 - A_1)D_{11} + (1 + A_1)D_{10}\right], \tag{B.12}$$

one obtains as a facet defining inequality for this scenario

$$\langle A(B - B') - A'(B + B')\rangle \leq 2. \tag{B.13}$$

C. Semi-definite programs

This appendix provides some general background on semi-definite programs and contains the details of the solution of the SDPs Eqs. (6.22) and (6.23) from Chapter 6.

C.1. Primal and dual of a semi-definite program

A semi-definite program (SDP) is a special class of convex optimisation problems (see e.g. Boyd and Vandenberghe, 2004). Let M_n denote the space of $n \times n$ matrices. A standard form SDP has linear constraints in form of equalities and a constraint of positive semi-definiteness for the variable $X \in M_n$:

$$\begin{aligned} p^\star = \text{maximise} \quad & \langle C, X \rangle \\ \text{subject to} \quad & \langle A_i, X \rangle = b_i, \quad i = 1, \ldots, m \\ & X \succeq 0, \end{aligned} \tag{C.1}$$

where $C, A_1, \ldots, A_m \in M_n$ and $\langle A, B \rangle = \mathrm{tr}(A^\top B)$ the inner product on M_n; note that every linear functional on M_n can be written as $X \mapsto \langle C, X \rangle$ for some $C \in M_n$.

One can write the above problem as an unconstrained one by introducing the Langragian

$$\mathcal{L}(X, \nu, Y) = \langle C, X \rangle + \sum_i \nu_i (b_i - \langle A_i, X \rangle) + \langle Y, X \rangle, \tag{C.2}$$

where $\nu \in \mathbb{R}^m$. Then the above maximisation can be expressed as

$$p^\star = \max_X \min_{Y \succeq 0} \mathcal{L}(X, \nu, Y). \tag{C.3}$$

The dual problem of Eq. (C.1) is then definied as

$$d^\star = \min_{\nu, Y \succeq 0} \max_X \mathcal{L}(X, \nu, Y), \tag{C.4}$$

where the minimax inequality immediately implies weak duality $p^\star \leq d^\star$. Making the constraints explicit, one has the following form for the dual problem

$$\begin{aligned} d^\star = \text{minimise} \quad & \nu^\top b \\ \text{subject to} \quad & \sum_i \nu_i A_i - C \succeq 0, \end{aligned} \tag{C.5}$$

which again is a SDP.

C.2. Solution of the SDPs from Chapter 6

Let us now turn to the SDPs encountered in Chapter 6. The primal problem Eq. (6.22) was given as

$$\begin{aligned} p^\star = \text{maximise} \quad & p \\ \text{subject to} \quad & \varrho \succeq 0, \\ & \operatorname{tr}_X \varrho = \varrho_{\text{red}}\,(p) \ \text{ for } X = A, B, C \\ & \varrho^{\top_X} \succeq 0 \text{ for } X = A, B, C, \end{aligned} \tag{C.6}$$

with the reduced state

$$\varrho_{\text{red}}\,(p) = \frac{2p}{3} \left|\psi^+\right\rangle\!\left\langle\psi^+\right| + \frac{p}{3} |00\rangle\langle 00| + \frac{1-p}{4} \mathbb{I}\,. \tag{C.7}$$

After introducing the dual variables $\mu = (\mu_A, \mu_B, \mu_C)$ and $\nu = (\nu_A, \nu_B, \nu_C)$ the Langragian for this problem is

$$\begin{aligned} &\mathcal{L}(\varrho, p, \mu, \nu) \\ &\quad = \frac{1}{4} \sum_X \langle \mu_X, \mathbb{I}\rangle + \frac{1}{4} p(1 + \sum_X \langle \mu_X, M\rangle) + \sum_X \langle \nu_X, \varrho^{\top_X}\rangle - \langle \mu_X, \operatorname{tr}_X \varrho\rangle \end{aligned} \tag{C.8}$$

$$\quad = \frac{1}{4} \sum_X \langle \mu_X, \mathbb{I}\rangle + \frac{1}{4} p(1 + \sum_X \langle \mu_X, M\rangle) + \sum_X \langle \nu_X^{\top_X} - \mathbb{I}_X \otimes \mu_X, \varrho\rangle, \tag{C.9}$$

where $\mu_X \in \mathbb{R}^{4\times 4}$, $\nu_X \in \mathbb{R}^{8\times 8}$ with $\nu_X \succeq 0$ for $X = A, B, C$ and

$$M = \frac{2}{3} \left|\psi^+\right\rangle\!\left\langle\psi^+\right| + \frac{1}{3} |00\rangle\langle 00| - \frac{1}{4} \mathbb{I}\,. \tag{C.10}$$

The expression $\mathbb{I}_X \otimes \mu_X$ denotes the operator that acts as the identity on the factor X and as μ_X on the rest.

Let us consider the maximisation of $\mathcal{L}(\varrho, p, \mu, \nu)$ over ϱ and p term by term. The first term is independent of both ϱ and p and corresponds to the objective function of the dual problem. For the second term we have

$$\max_{p\geq 0} \frac{1}{4} p(1 + \sum_X \langle \mu_X, M\rangle) = \begin{cases} 0 & \text{if } \sum_X \langle \mu_X, M\rangle \leq -1 \\ \infty & \text{otherwise,} \end{cases} \tag{C.11}$$

whereas the last term gives

$$\max_{\substack{\varrho\succeq 0 \\ \operatorname{tr}\varrho = 1}} \sum_X \langle \nu_X^{\top_X} - \mathbb{I}_X \otimes \mu_X, \varrho\rangle = \begin{cases} 0 & \text{if } \sum_X (\mathbb{I}_X \otimes \mu_X - \nu_X^{\top_X}) \succeq 0 \\ \infty & \text{otherwise.} \end{cases} \tag{C.12}$$

Thus, the dual problem of Eq. (C.6) can be expressed as

$$\begin{aligned} d^\star = \text{minimise} \quad & \frac{1}{4}\operatorname{tr}(\mu_A + \mu_B + \mu_C) \\ \text{subject to} \quad & \nu_X \succeq 0, \\ & \operatorname{tr}\left[(\mu_A + \mu_B + \mu_C)M\right] \leq -1, \\ & \sum_X \mathbb{I}_X \otimes \mu_X - \nu_X^{\top_X} \succeq 0. \end{aligned} \tag{C.13}$$

From weak duality one always has $d^\star \geq p^\star$. Every feasible point for the primal problem gives a lower bound $p' \leq p^\star$ and every dual feasible point gives an upper bound $d' \geq d^\star$. In what follows we provide a choice of the variables ϱ, μ_X, ν_X that satisfy all the constraints of Eq. (C.6) and Eq. (C.13) while yielding the same bounds $d' = p'$, we thus have strong duality $p^\star = d^\star$, where

$$p^\star = d^\star = \frac{3}{2 + \sqrt{17}}. \tag{C.14}$$

The density matrix ϱ that attains the maximum of Eq. (C.6) is

$$\begin{aligned} \varrho =& \frac{p^\star}{2}\left(|W\rangle\langle W| + |\overline{W}\rangle\langle\overline{W}|\right) + \frac{3(1-p^\star)}{4}\sigma \\ &+ \frac{p^\star}{6}\,|000\rangle\langle 000| + \frac{3-5p^\star}{12}\,|111\rangle\langle 111| \end{aligned} \tag{C.15}$$

with

$$|\overline{W}\rangle = \frac{1}{\sqrt{3}}\left(|011\rangle + |101\rangle + |110\rangle\right), \tag{C.16}$$

$$\sigma = \frac{1}{3}\left(|001\rangle\langle 001| + |010\rangle\langle 010| + |100\rangle\langle 100|\right). \tag{C.17}$$

The optimum of Eq. (C.13) can be attained with $\mu_A = \mu_B = \mu_C = \mu$, where, with respect to the basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$,

$$\mu = \begin{pmatrix} a & 0 & 0 & 0 \\ 0 & b & c & 0 \\ 0 & c & b & 0 \\ 0 & 0 & 0 & d \end{pmatrix} \tag{C.18}$$

and

$$\begin{aligned} a &= \frac{p^\star}{2}(1 + \frac{5}{3\sqrt{17}}), \quad b = \frac{p^\star}{12}(1 - \sqrt{17}), \\ c &= \frac{p^\star}{6}(1 + \frac{11}{\sqrt{17}}), \quad d = p^\star(\frac{2}{3} + \frac{2}{\sqrt{17}}). \end{aligned} \tag{C.19}$$

In the basis $\{|000\rangle, |001\rangle, |010\rangle, |011\rangle, |100\rangle, |101\rangle, |110\rangle, |111\rangle\}$ the matrix ν_A reads

$$\nu_A = \begin{pmatrix} a & 0 & 0 & 0 & 0 & a & a & 0 \\ 0 & b & b & 0 & 0 & 0 & 0 & c \\ 0 & b & b & 0 & 0 & 0 & 0 & c \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ a & 0 & 0 & 0 & 0 & d & e & 0 \\ a & 0 & 0 & 0 & 0 & e & d & 0 \\ 0 & c & c & 0 & 0 & 0 & 0 & f \end{pmatrix}, \tag{C.20}$$

$$\begin{aligned} a &= \frac{p^\star}{4}(1 + \frac{5}{3\sqrt{17}}), \quad b = -p^\star(\frac{1}{3} - \frac{1}{\sqrt{17}}), \\ c &= p^\star \frac{4}{3\sqrt{17}}, \quad d = -\frac{p^\star}{2}(\frac{3}{5} - \frac{1}{3\sqrt{17}}), \\ e &= \frac{p^\star}{4}(\frac{1}{5} - \frac{7}{3\sqrt{17}}), \quad f = -2p^\star(\frac{1}{3} + \frac{1}{\sqrt{17}}). \end{aligned} \tag{C.21}$$

The matrices ν_B and ν_C can be obtained from ν_A by interchanging the role of A with B or C respectively.

Bibliography

Acín, A., T. Durt, N. Gisin, and J. I. Latorre, 2002. Quantum nonlocality in two three-level systems. *Phys. Rev. A*, 65(5), 052325.

Acín, Antonio, Nicolas Brunner, Nicolas Gisin, Serge Massar, Stefano Pironio, and Valerio Scarani, 2007. Device-Independent Security of Quantum Cryptography against Collective Attacks. *Phys. Rev. Lett.*, 98(23), 230501.

Ahanj, Ali, Samir Kunkri, Ashutosh Rai, Ramij Rahaman, and Pramod S. Joag, 2010. Bound on Hardy's nonlocality from the principle of information causality. *Phys. Rev. A*, 81(3), 032103.

Allcock, Jonathan, Nicolas Brunner, Marcin Pawlowski, and Valerio Scarani, 2009. Recovering part of the boundary between quantum and nonquantum correlations from information causality. *Phys. Rev. A*, 80(4), 040103.

Almeida, Mafalda L., Jean-Daniel Bancal, Nicolas Brunner, Antonio Acín, Nicolas Gisin, and Stefano Pironio, 2010. Guess Your Neighbor's Input: A Multipartite Nonlocal Game with No Quantum Advantage. *Phys. Rev. Lett.*, 104(23), 230404.

Aspect, Alain, Jean Dalibard, and Gérard Roger, 1982. Experimental Test of Bell's Inequalities Using Time- Varying Analyzers. *Phys. Rev. Lett.*, 49(25), 1804.

Barrett, J., 2002. Nonsequential positive-operator-valued measurements on entangled mixed states do not always violate a Bell inequality. *Phys. Rev. A*, 65, 042302.

Barrett, Jonathan, Lucien Hardy, and Adrian Kent, 2005. No Signaling and Quantum Key Distribution. *Phys. Rev. Lett.*, 95(1), 010503.

Barrett, Jonathan and Stefano Pironio, 2005. Popescu-Rohrlich Correlations as a Unit of Nonlocality. *Phys. Rev. Lett.*, 95(14), 140401.

Barrett, Jonathan, Stefano Pironio, Jean-Daniel Bancal, and Nicolas Gisin, 2011. The definition of multipartite nonlocality. arXiv:1112.2626 [quant-ph].

Bell, John, 1964. On the Einstein Podolsky Rosen Paradox. *Physics*, 1, 195.

Bennett, C. H. and G. Brassard, 1984. Quantum Cryptography: Public key distribution and coin tossing. In *Proceedings of the IEEE International Conference on Computers, Systems, and Signal Processing*, Bangalore, India, 175.

Boyd, Stephen and Lieven Vandenberghe, 2004. *Convex Optimization*. New York, NY, USA: Cambridge University Press.

Branciard, Cyril, Denis Rosset, Nicolas Gisin, and Stefano Pironio, 2012. Bilocal versus nonbilocal correlations in entanglement-swapping experiments. *Phys. Rev. A*, 85(3), 032119.

Brassard, Gilles, Harry Buhrman, Noah Linden, André Allan Méthot, Alain Tapp, and Falk Unger, 2006. Limit on Nonlocality in Any World in Which Communication Complexity Is Not Trivial. *Phys. Rev. Lett.*, 96(25), 250401.

Brunner, Nicolas and Paul Skrzypczyk, 2009. Nonlocality Distillation and Postquantum Theories with Trivial Communication Complexity. *Phys. Rev. Lett.*, 102(16), 160403.

Cavalcanti, Daniel, Alejo Salles, and Valerio Scarani, 2010. Macroscopically local correlations can violate information causality. *Nat. Commun.*, 1, 136.

Clauser, John F., Michael A. Horne, Abner Shimony, and Richard A. Holt, 1969. Proposed Experiment to Test Local Hidden-Variable Theories. *Phys. Rev. Lett.*, 23(15), 880.

Clifton, Rob, Jeffrey Bub, and Hans Halvorson, 2003. Characterizing Quantum Theory in Terms of Information-Theoretic Constraints. *Foundations of Physics*, 33(11), 1561.

Coffman, Valerie, Joydip Kundu, and William K. Wootters, 2000. Distributed entanglement. *Phys. Rev. A*, 61(5), 052306.

Colbeck, Roger and Adrian Kent, 2011. Private randomness expansion with untrusted devices. *Journal of Physics A: Mathematical and Theoretical*, 44(9), 095305.

van Dam, Wim, 2000. *Nonlocality and communication complexity*. Ph.D. thesis, University of Oxford.

Einstein, A., B. Podolsky, and N. Rosen, 1935. Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? *Phys. Rev.*, 47, 777.

Ekert, A. K., 1991. Quantum cryptography based on Bell's theorem. *Phys. Rev. Lett.*, 67, 661.

Fine, A., 1982. Hidden Variables, Joint Probability, and the Bell Inequalities. *Phys. Rev. Lett.*, 48, 291.

Fritz, Tobias, 2010. On the existence of quantum representations for two dichotomic measurements. *J. Math. Phys.*, 51(5), 052103.

Fritz, Tobias, 2012. Beyond Bell's theorem: correlation scenarios. *New Journal of Physics*, 14(10), 103001.

Fritz, Tobias, Anthony Leverrier, and Ana Belén Sainz, 2012a. A Combinatorial Approach to Nonlocality and Contextuality. arXiv:1212.4084 [quant-ph].

Fritz, Tobias, Belén Sainz, Remigiusz Augusiak, Jonatan Bohr Brask, Rafael Chaves, Anthony Leverrier, and Antonio Acín, 2012b. Local Orthogonality: a multipartite principle for correlations. arXiv:1210.3018 [quant-ph].

Gallego, Rodrigo, Lars Erik Würflinger, Antonio Acín, and Miguel Navascués, 2012. Operational Framework for Nonlocality. *Phys. Rev. Lett.*, 109(7), 070401.

Gallego, Rodrigo, Lars Erik Würflinger, Antonio Acín, and Miguel Navascués, 2011. Quantum Correlations Require Multipartite Information Principles. *Phys. Rev. Lett.*, 107(21), 210403.

Gisin, N., 1991. Bell's inequality holds for all non-product states. *Phys. Lett. A*, 154, 201.

Gisin, N., 1996. Hidden quantum nonlocality revealed by local filters. *Physics Letters A*, 210(3), 151.

Gittsovich, Oleg, Philipp Hyllus, and Otfried Gühne, 2010. Multiparticle covariance matrices and the impossibility of detecting graph-state entanglement with two-particle correlations. *Phys. Rev. A*, 82(3), 032306.

Hall, William, 2007. Compatibility of subsystem states and convex geometry. *Phys. Rev. A*, 75(3), 032102.

Higuchi, A., A. Sudbery, and J. Szulc, 2003. One-Qubit Reduced States of a Pure Many-Qubit State: Polygon Inequalities. *Phys. Rev. Lett.*, 90(10), 107902.

Horodecki, R., P. Horodecki, and M. Horodecki, 1995. Violating Bell inequality by mixed states: necessary and sufficient condition. *Phys. Lett. A*, 200(5), 340.

Horodecki, Ryszard, Paweł, Horodecki, Michał, Horodecki, and Karol Horodecki, 2009. Quantum entanglement. *Rev. Mod. Phys.*, 81(2), 865.

Klyachko, Alexander, 2004. Quantum marginal problem and representations of the symmetric group. arXiv:quant-ph/0409113.

Koashi, Masato and Andreas Winter, 2004. Monogamy of quantum entanglement and other correlations. *Phys. Rev. A*, 69(2), 022309.

Linden, N., S. Popescu, and W. K. Wootters, 2002. Almost Every Pure State of Three Qubits Is Completely Determined by Its Two-Particle Reduced Density Matrices. *Phys. Rev. Lett.*, 89(20), 207901.

Löfberg, J, 2004. YALMIP : A Toolbox for Modeling and Optimization in MATLAB. In *Proceedings of the CACSD Conference*, Taipei, Taiwan.

Masanes, Lluis, Stefano Pironio, and Antonio Acin, 2011. Secure device-independent quantum key distribution with causally independent measurement devices. *Nat. Commun.*, 2, 238.

Methot, A. A. and V. Scarani, 2007. An anomaly of non-locality. *Quantum Information and Computation*, 7, 157.

Navascués, Miguel, Stefano Pironio, and Antonio Acín, 2007. Bounding the Set of Quantum Correlations. *Phys. Rev. Lett.*, 98(1), 010401.

Navascués, Miguel, Stefano Pironio, and Antonio Acín, 2008. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New Journal of Physics*, 10(7), 073013.

Osterloh, A., Luigi Amico, G. Falci, and Rosario Fazio, 2002. Scaling of entanglement close to a quantum phase transition. *Nature*, 416(6881), 608.

Pawlowski, Marcin, Tomasz Paterek, Dagomir Kaszlikowski, Valerio Scarani, Andreas Winter, and Marek Zukowski, 2009. Information causality as a physical principle. *Nature*, 461(7267), 1101.

Pearl, J, 2009. *Causality: Models, Reasoning, and Inference.* Second edition, Cambridge University Press.

Peres, Asher, 1996. Separability Criterion for Density Matrices. *Phys. Rev. Lett.*, 77, 1413.

Pironio, S., 2005. Lifting Bell inequalities. *Journal of Mathematical Physics*, 46, 062112.

Pironio, S., A. Acín, S. Massar, A. Boyer de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, and C. Monroe, 2010. Random numbers certified by Bell's theorems. *Nature*, 464(7291), 1021.

Pironio, Stefano, Jean-Daniel Bancal, and Valerio Scarani, 2011. Extremal correlations of the tripartite no-signaling polytope. *Journal of Physics A: Mathematical and Theoretical*, 44(6), 065303.

Pitowsky, Itamar, 1989. *Quantum Probability, Quantum Logic*, volume 321 of *Lecture Notes in Physics*. Heidelberg: Springer.

Popescu, Sandu, 1995. Bell's Inequalities and Density Matrices: Revealing "Hidden" Nonlocality. *Phys. Rev. Lett.*, 74(14), 2619.

Popescu, Sandu and Daniel Rohrlich, 1994. Quantum nonlocality as an axiom. *Foundations of Physics*, 24(3), 379.

Popescu, Sandu, Anthony J. Short, and Andreas Winter, 2006. Entanglement and the foundations of statistical mechanics. *Nat. Phys.*, 2(11), 754.

Schrödinger, E., 1935. Die gegenwärtige Situation in der Quantenmechanik. *Naturwissenschaften*, 23, 807.

Shor, P. W., 1994. Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer. *SIAM J. Sci. Statist. Comput.*, 26, 1484.

Spirtes, P., N. Glymour, and R. Scheienes, 2001. *Causation, Prediction, and Search*. Second edition, The MIT Press.

Svetlichny, G., 1987. Distinguishing three-body from two-body nonseparability by a Bell-type inequality. *Phys. Rev. D*, 35, 3066.

Teufel, Stefan, Karin Berndl, Detlef Dürr, Sheldon Goldstein, and Nino Zanghì, 1997. Locality and causality in hidden-variables models of quantum theory. *Phys. Rev. A*, 56, 1217.

Tsirelson, B., 1983. *Lett. Math. Phys.*, 4, 93.

Tóth, Géza, Christian Knapp, Otfried Gühne, and Hans J. Briegel, 2007. Optimal Spin Squeezing Inequalities Detect Bound Entanglement in Spin Models. *Phys. Rev. Lett.*, 99(25), 250405.

Tóth, Géza, Christian Knapp, Otfried Gühne, and Hans J. Briegel, 2009. Spin squeezing and entanglement. *Phys. Rev. A*, 79(4), 042334.

Valentini, Antony, 2002. Signal-locality in hidden-variables theories. *Phys. Lett. A*, 297(5–6), 273.

Werner, R. F. and M. M. Wolf, 2001. Bell inequalities and entanglement. *Quantum Information and Computation*, 1, 1.

Werner, Reinhard F., 1989. Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model. *Phys. Rev. A*, 40, 4277.

Wolf, Michael M., Toby S. Cubitt, and David Pérez-García, 2011. Are problems in Quantum Information Theory (un)decidable? arXiv:1111.5425 [quant-ph].

Würflinger, Lars Erik, Jean-Daniel Bancal, Antonio Acín, Nicolas Gisin, and Tamás Vértesi, 2012. Nonlocal multipartite correlations from local marginal probabilities. *Phys. Rev. A*, 86(3), 032117.

Yang, Tzyh Haur, Daniel Cavalcanti, Mafalda L Almeida, Colin Teo, and Valerio Scarani, 2012. Information-causality and extremal tripartite correlations. *New Journal of Physics*, 14(1), 013061.

Yao, Andrew Chi-Chih, 1979. Some complexity questions related to distributive computing (Preliminary Report). In *Proceedings of the eleventh annual ACM symposium on Theory of computing*, STOC '79, New York, NY, USA: ACM, 209–213.