Contextuality as a precondition for entanglement

Martin Plávala and Otfried Gühne

Naturwissenschaftlich-Technische Fakultät, Universität Siegen, 57068 Siegen, Germany

(Dated: September 22, 2022)

Quantum theory features several phenomena which can be considered as resources for information processing tasks. Some of these effects, such as entanglement, arise in a non-local scenario, where a quantum state is distributed between different parties. Other phenomena, such as contextuality, can be observed, if quantum states are prepared and then subjected to sequences of measurements. Here we provide an intimate connection between different resources by proving that entanglement in a non-local scenario can only arise if there is preparation & measurement contextuality in a sequential scenario derived from the non-local one by remote state preparation. Moreover, the robust absence of entanglement implies the absence of contextuality. As a direct consequence, our result allows to translate any inequality for testing preparation & measurement contextuality into an entanglement test; in addition, entanglement witnesses can be used to obtain novel contextuality inequalities.

Introduction.— Quantum information science bears the promise to lead to novel ways of information processing, which are superior to classical methods. This begs the question, which quantum phenomena are responsible for the quantum advantage and which resources are needed to overcome classical limits. There are two main scenarios where genuine quantum effects are studied. First, in the nonlocal scenario (NLS), two parties, Alice and Bob, share a bipartite quantum state $\rho_{AB}$ and perform different measurements on it. This leads to a joint probability distribution for the possible outcomes. Second, in the sequential scenario (SQS), Alice prepares some quantum state $\sigma$, transmits the quantum state to Bob which performs a measurement. Clearly, these scenarios are connected: In the NLS Alice and Bob can, using classical communication, postselect on the outcome of Alice’s measurement, so that Alice remotely prepares the state $\sigma$ for Bob, see also Fig. 1.

In both scenarios, several notions of classicality are known and have been identified as resources for special tasks. For the NLS a major example is entanglement which arises if the quantum state cannot be generated by local operations and classical communication [1, 2]. Entanglement has been identified as a resource for tasks like quantum key distribution [3] or quantum metrology [4, 5]. Other examples of non-classicality in the NLS are quantum steering and [6] and Bell nonlocality [7].

For the SQS a major notion of non-classicality is based on quantum contextuality [8, 9]. For preparation noncontextuality one asks whether there is a hidden variable (HV) model for Alice’s preparations obeying some assumptions, while measurement noncontextuality refers to the same question for Bob’s measurements. Often, the underlying notions of classicality are combined to preperation and measurement (P&M) contextuality, sometimes also called simplex embeddability [10–13]. Effects of contextuality can be viewed as resources in various tasks, such as quantum state discrimination [14] and parity-oblivious multiplexing [15].

Are there any connections between quantum resources arising in the nonlocal and the sequential scenario? This is a key question for understanding the quantum advantage in information processing. For quantum key distribution it was already observed some time ago that prepare and measure schemes (like the BB84 protocol) can be mapped to entanglement-based schemes, which allows for a common security analysis of both scenarios based on entanglement theory [3, 16]. More recently, the notion of remote state preparation was used to show that steerability of a quantum state $\rho_{AB}$ corresponds to preparation noncontextuality [17], while steerability of an assemblage corresponds to measurement noncontextuality [18], see also Ref. [12] for a discussion.

In this paper we show that entanglement in the NLS corresponds to P&M contextuality in the SQS. We use the fact that any bipartite quantum state gives rise to some sequential scenario by using a kind of remote state preparation [19, 20]. Then, P&M contextuality in this SQS is precondition of entanglement of the bipartite state. Our results imply that one can map noncontextu-
ality inequalities to entanglement witnesses and that one can use classes of entanglement witnesses to obtain noncontextuality inequalities. Our research was motivated by recent findings on a connection between noncontextuality and the mathematical notion of generalized separability of the identity map [21].

Entanglement and remote preparations.— Assume that two remote parties, Alice and Bob, share a bipartite quantum state \( \rho_{AB} \). We then say that the bipartite state \( \rho_{AB} \) is entangled if it cannot be prepared using local operations and classical communication [22]. This is the same as requiring that the state \( \rho_{AB} \) is not separable, i.e., there is no decomposition of the form \( \rho_{AB} = \sum_i p_i \sigma_i^A \otimes \sigma_i^B \), where the \( p_i \) form a probability distribution and \( \sigma_i^A \) and \( \sigma_i^B \) are some states of Alice's and Bob's system.

Given a bipartite state \( \rho_{AB} \) Bob can apply a measurement to his part of the system and announce the outcome, which results in remotely preparing the state \( \text{Tr}_B[(I_A \otimes E_B)\rho_{AB}] \) for Alice. While a single remotely prepared state does not capture the properties of the bipartite state \( \rho_{AB} \) shared between Alice and Bob, it is intuitive that the set of all possible remotely preparable states should have some properties based on whether the shared bipartite state \( \rho_{AB} \) is entangled or not. In order to investigate this, we denote by \( \Lambda_A(\rho_{AB}) \) the set of all possible states that Bob can remotely prepare for Alice using the shared bipartite state \( \rho_{AB} \). Mathematically \( \Lambda_A(\rho_{AB}) \) is defined as

\[
\Lambda_A(\rho_{AB}) = \{ \sigma_A \in \mathcal{D}(\mathcal{H}_A) : E_B \geq 0, \sigma_A = \text{Tr}_B[(I_A \otimes E_B)\rho_{AB}] \},
\]

where \( \mathcal{H}_A \) is the Hilbert space corresponding to Alice's system, \( \mathcal{D}(\mathcal{H}_A) \) is the set of density matrices on \( \mathcal{H}_A \), and \( E_B \geq 0 \) means that \( E_B \) is a positive semidefinite operator. \( \Lambda_B(\rho_{AB}) \) is defined analogically. Note that these sets of quantum states play a role in recent approaches to tackle the problem of quantum steering [20, 23].

Contextuality.— There are several notions of contextuality, here we are interested in so-called preparation & measurement (P&M) noncontextual models in the sense of Spekkens [8, 9]. In this approach one considers a set of preparations and a set of measurements. Two preparations are operationally equivalent if, for all measurements, they give rise to the same probability distribution of outcomes. Analogously, two measurements are operationally equivalent if they result in the same probability distribution for any available state preparation.

In a HV model for this scenario, the assumption of preparation noncontextuality states that two equivalent preparations give rise to the same probability distribution over the HV. Similarly, measurement noncontextuality assumes that equivalent measurements are described by the same response functions. In addition, for a HV model one makes the general assumption of convex linearity. This means that if one chooses randomly one of two preparations, then the resulting probability distribution of the HV is the mixture of the two distributions of the preparations. An analogous assumption is made on random choices of measurements.

In quantum mechanics, state preparations are described by density matrices and measurement probabilities are computed by effects. Thus, a quantum system has a P&M noncontextual HV model if for every density matrix \( \rho \in K \) (where \( K \) is a convex subset of density matrices that corresponds to the prepareable states) and for every POVM \( M = \{ M_a \} \), \( 0 \leq M_a \leq 1 \), \( \sum_a M_a = 1 \), we have

\[
\text{Tr}(\rho M_a) = \sum_\lambda p(\lambda|\rho) p(a|\lambda, M)
\]

where \( p(\lambda|\rho) \) is a linear function of \( \rho \) and \( p(a|\lambda, M) \) is a linear function of \( M_a \), see [17] for a detailed explanation.

For our purposes, it is useful to rewrite Eq. (2). Since \( p(\lambda|\rho) \) is a linear function of \( \rho \) it follows that there are operators \( N_\lambda \) such that \( p(\lambda|\rho) = \text{Tr}(\rho N_\lambda) \) that satisfy the positivity and normalization conditions:

\[
\text{Tr}(\rho N_\lambda) \geq 0 \quad \text{and} \quad \sum_\lambda \text{Tr}(\rho N_\lambda) = 1,
\]

for all \( \rho \in K \). Note that this does not imply that \( N_\lambda \) is a POVM. In fact, whenever \( K \) is a strict subset of the density matrices, then according to the hyperplane separation theorem [24] there is some \( N_\lambda \) that satisfies the positivity condition in Eq. (3) that is not a positive semidefinite operator. Analogically, we require that \( p(a|\lambda, M) \) is a linear function of \( M_a \) so we have \( p(a|\lambda, M) = \text{Tr}(\omega_\lambda M_a) \), where \( \omega_\lambda \in \mathcal{D}(\mathcal{H}) \). Putting everything together, we get that there exists preparation & measurement noncontextual HV model for states in \( K \) if for every \( \rho \in K \) and every POVM \( M = \{ M_a \} \) we have

\[
\text{Tr}(\rho M_a) = \sum_\lambda \text{Tr}(\rho N_\lambda) \text{Tr}(\omega_\lambda M_a).
\]

Main Results.— We can directly formulate our first main result:

**Theorem 1.** Let \( \rho_{AB} \) be a bipartite quantum state and assume that there exists a P&M noncontextual model for the set of states \( \Lambda_A(\rho_{AB}) \) and all possible measurements. Then, \( \rho_{AB} \) is separable.

The proof is given in Appendix A. The underlying idea is that in the P&M noncontextual model from Eq. (4) the term \( \text{Tr}(\omega_\lambda M_a) \) can be interpreted as measurement on Alice's side of a separable decomposition \( \rho_{AB} = \sum_\lambda \omega_\lambda \otimes K_\lambda \), where Bob's parts \( K_\lambda \) in the decomposition are given by \( K_\lambda = \text{Tr}_A[(N_\lambda \otimes I_B)\rho_{AB}] \).

In the following theorem we will prove that a separable state \( \rho_{AB} = \sum_\lambda \omega_\lambda \otimes K_\lambda \) yields remotely preparable set \( \Lambda_A(\rho_{AB}) \) with a preparation & measurement noncontextual model if the operators \( K_\lambda \) in the separable decomposition can be chosen such that they belong to the linear
hull of $\Lambda_B(\rho_{AB})$. We will get rid of this condition later by considering robust remote preparations.

**Theorem 2.** Let $\rho_{AB}$ be a separable bipartite quantum state with the decomposition $\rho_{AB} = \sum_\lambda \omega_\lambda \otimes K_\lambda$, where $\omega_\lambda \geq 0$, $\text{Tr}(\omega_\lambda) = 1$ and $K_\lambda \geq 0$. Assume that $K_\lambda$ belongs to the linear hull of $\Lambda_B(\rho_{AB})$ for all $\lambda$. Then there exists P&\text{M noncontextual model for } \Lambda_A(\rho_{AB}).$

The proof can be found in Appendix B. The following example shows that Theorem 2 cannot be directly extended to all states. Consider the separable qubit-ququart state

$$\rho_{AB} = \frac{1}{4}(|0\rangle\langle 0| \otimes |00\rangle\langle 00| + |1\rangle\langle 1| \otimes |01\rangle\langle 01| + |+\rangle\langle +| \otimes |10\rangle\langle 10| + |-\rangle\langle -| \otimes |11\rangle\langle 11|). \tag{5}$$

This does not meet the condition in Theorem 2; at least for the decomposition in Eq. (5) this is obvious: $\Lambda_B(\rho_{AB})$ is three-dimensional, while there are four linearly independent $K_\lambda$. On the other hand, we have $\Lambda_A(\rho_{AB}) = \text{conv}(\{0|0\rangle\langle 0|, 1|1\rangle\langle 1|, |+\rangle\langle +|, |-\rangle\langle -|\})$, and we will show below that a preparation & measurement noncontextual model for $\Lambda_A(\rho_{AB})$ does not exist using violation of the contextuality inequality that we will derive in Proposition 5, see Eq. (15) below.

Clearly, the linear hull condition in the Theorem 2 is met, if the set $\Lambda_B(\rho_{AB})$ spans the entire operator space. Not surprisingly, this is the generic case, and one can get rid of the pathological cases by considering robust remote preparations, i.e., by adding an infinitesimal amount of random separable noise. This idea is summarized in the following theorem, the proof can be found in the Appendix C.

**Theorem 3.** Let $\rho_{AB}$ be a separable quantum state. Then for almost all separable quantum states $\tau_{AB}$ there is a $\delta(\tau_{AB}) > 0$ depending on $\tau_{AB}$, such that for every $\epsilon \in (0, \delta]$ there exists P&\text{M noncontextual model for } $\Lambda_A[(1-\epsilon)\rho_{AB} + \epsilon\tau_{AB}]$.

**Mapping noncontextuality inequalities to entanglement witnesses.**— Using the results of Theorem 3 one can obtain entanglement witnesses from noncontextuality inequalities. The only caveat is that the noncontextuality inequalities must be formulated in terms of unnormalized states, this is necessary to account for the fact that $\text{Tr}_B((\mathbb{1}_A \otimes \mathbb{E}_B)\rho_{AB})$ is not normalized for $\mathbb{E}_B \geq 0$. This is because the set of $\mathbb{E}_B$ such that $\text{Tr}_B((\mathbb{1}_A \otimes \mathbb{E}_B)\rho_{AB}) = 1$ in general depends on $\rho_{AB}$.

We will demonstrate the method using the noncontextuality inequality presented in Ref. [25]. Let $K$ be the set of allowed preparations. By cone($K$) we denote the set of all unnormalized allowed preparations, that is all operators of the form $\mu \sigma$, where $\mu \in \mathbb{R}, \mu \geq 0$ and $\sigma \in K$. Let $\sigma_{t,b} \in \text{cone}(K)$ for $t \in \{1, 2, 3\}$ and $b \in \{0, 1\}$ be such that $\sigma_* = \frac{1}{2}(\sigma_{t,0} + \sigma_{t,1})$ is the same for all $t \in \{1, 2, 3\}$ and let $M_{t,b}$ be positive operators, $M_{t,b} \geq 0$, such that $\frac{1}{3} \sum_{t=1}^3 M_{t,b} = \frac{1}{2}$ and $M_{t,0} + M_{t,1} = 1$. In other words, $M_{t,b}$ are three binary POVMs such that their uniform mixture corresponds to the random coin toss. Then the unnormalized version of noncontextuality inequality from Ref. [25] is as follows: If there is a P&M noncontextual model for $K$, then we have

$$\sum_{t=1}^3 \sum_{b=0}^1 \text{Tr}(\sigma_{t,b} M_{t,b}) \leq 5 \text{Tr}(\sigma_*). \tag{6}$$

See Appendix D for the proof of this modified noncontextuality inequality. Using this inequality we obtain the following entanglement witness:

**Proposition 4.** Let $\rho_{AB}$ be a separable quantum state. Let $E_{t,b}$ and $M_{t,b}$ be positive operators, $E_{t,b} \geq 0$ and $M_{t,b} \geq 0$, such that $E_* = \frac{1}{2}(E_{t,0} + E_{t,1})$, $\frac{1}{3} \sum_{t=1}^3 M_{t,b} = \frac{1}{2}$ and $\mathbb{1}_A = M_{t,0} + M_{t,1}$ for all $t \in \{1, 2, 3\}$ and $b \in \{0, 1\}$. Then

$$\sum_{t=1}^3 \sum_{b=0}^1 \text{Tr}[(M_{t,b} \otimes E_{t,b})\rho_{AB}] \leq 5 \text{Tr}[(\mathbb{1}_A \otimes E_*)\rho_{AB}]. \tag{7}$$

Moreover, there is an entangled state $\rho_{AB}$ that violates Eq. (7) for suitable choice of the operators $E_{t,b}$ and $M_{t,b}$.

The proof follows from Theorem 3, since if $\rho_{AB}$ is separable, then Eq. (7) is satisfied for all $\Lambda_A[(1-\epsilon)\rho_{AB} + \epsilon\tau_{AB}]$ for all $\epsilon \in (0, \delta]$. This is then used to construct the corresponding entanglement witness. The full proof can be found in Appendix D.

Being more concrete, one can write down an explicit entanglement witness from observables leading to a violation of the noncontextuality inequality [9]. We consider qubit systems and the Pauli matrices $\sigma_x, \sigma_y, \sigma_z$. Then we define the effects

$$E_{1,0} = \mathbb{1} + \sigma_z, \quad E_{1,1} = \mathbb{1} - \sigma_z, \quad E_{2,0} = \mathbb{1} + \frac{\sqrt{3}}{2} \sigma_x - \frac{1}{2} \sigma_z, \quad E_{x,1} = \mathbb{1} - \frac{\sqrt{3}}{2} \sigma_x + \frac{1}{2} \sigma_z, \tag{8}$$

$$E_{3,0} = \mathbb{1} - \frac{\sqrt{3}}{2} \sigma_x - \frac{1}{2} \sigma_z, \quad E_{3,1} = \mathbb{1} + \frac{\sqrt{3}}{2} \sigma_x + \frac{1}{2} \sigma_z, \tag{9}$$

and $M_{t,b} = \frac{1}{2} E_{t,b}$. From this we obtain the entanglement witness $W = 2 \mathbb{1}_A \otimes \mathbb{1}_B - \frac{3}{2}(\sigma_x \otimes \sigma_x + \sigma_z \otimes \sigma_z)$, so for every separable state $\rho_{AB}$ we have

$$\text{Tr}[(\sigma_x \otimes \sigma_x)\rho_{AB}] + \text{Tr}[(\sigma_z \otimes \sigma_z)\rho_{AB}] \leq \frac{4}{3}. \tag{11}$$

which is a weakened version of the well-known witness $\text{Tr}[(\sigma_x \otimes \sigma_x)\rho_{AB}] + \text{Tr}[(\sigma_z \otimes \sigma_z)\rho_{AB}] \leq 1$ [26–28]; still, the inequality (11) is violated by the maximally entangled state.
Mapping entanglement witnesses to noncontextuality inequalities.— This direction is not so straightforward, as we cannot use just a single entanglement witness, but we must map a whole class of entanglement witnesses to get a class of noncontextuality inequalities. We will proceed with an example that showcases this.

We will use a class of entanglement witnesses that comes from the Clauser-Horne-Shimony-Holt (CHSH) inequality [29, 30]. Let $A_i$ and $B_i$ for $i \in \{1, 2\}$ be observables such that $-\mathbb{1}_A \leq A_i \leq \mathbb{1}_A$ and $-\mathbb{1}_B \leq B_i \leq \mathbb{1}_B$ and let $\rho_{AB}$ be a separable state. Then we have

$$
\text{Tr}\{[A_1 \otimes (B_1 + B_2) + A_2 \otimes (B_1 - B_2)]\rho_{AB}\} \leq 2.
$$

In order to obtain a noncontextuality inequality proceed as follows. Let $B_{i+}$ and $B_{i-}$ be the positive and negative parts of $B_i$, respectively, and denote $\sigma_{i\pm} = \text{Tr}_B[(I_A \otimes B_{i\pm})\rho_{AB}]$. Then we have

$$
2 \geq \text{Tr}\{[A_1 \otimes (B_1 + B_2) + A_2 \otimes (B_1 - B_2)]\rho_{AB}\} = \text{Tr}\{(A_1 + A_2)(\sigma_{1+} - \sigma_{1-})\} + \text{Tr}\{(A_1 - A_2)(\sigma_{2+} - \sigma_{2-})\},
$$

which bears already some formal similarity to the noncontextuality inequality from above. Moreover, we have $B_{i+} = B_{i+} - B_{i-}$ and $|B_i| = B_{i+} + B_{i-}$. From $-\mathbb{1}_B \leq B_i \leq \mathbb{1}_B$ it follows that the eigenvalues of $B_i$ are from the interval $[-1, 1]$ and so we also have $|B_i| \leq \mathbb{1}_B$.

Let us define $\sigma_0 = \text{Tr}_B[(I_A \otimes (B_0 - |B_i|))\rho_{AB}]$ and $\sigma_* = \text{Tr}_B(\rho_{AB})$. Then we have $\sigma_{1+} + \sigma_{1-} + \sigma_{10} = \sigma_* = \sigma_{2+} + \sigma_{2-} + \sigma_{20}$, which is going to play a crucial role in the formulation of the noncontextuality inequality. We obtain:

**Proposition 5.** Let $K$ be a set of allowed preparations. Let $\sigma_* \in K$ and let $i \in \{1, 2\}$, let $\sigma_{1+}, \sigma_{1-}, \sigma_{10} \in \text{cone}(K)$ be subnormalized preparations such that

$$
\sigma_{1+} + \sigma_{1-} + \sigma_{10} = \sigma_* = \sigma_{2+} + \sigma_{2-} + \sigma_{20}.
$$

Let $A_i$ be observables such that $-\mathbb{1} \leq A_i \leq \mathbb{1}$ for all $i \in \{1, 2\}$. If there is a P&M contextual model for $K$, then

$$
\text{Tr}\{(A_1 + A_2)(\sigma_{1+} - \sigma_{1-})\} + \text{Tr}\{(A_1 - A_2)(\sigma_{2+} - \sigma_{2-})\} \leq 2.
$$

(14)

The proof of Proposition 5 is given in Appendix E. The proof is significantly different from the proof of Proposition 4: There, we showed that Eq. (7) is an entanglement witness as a result of Eq. (6) being noncontextuality inequality. In the proof of Proposition 5 we use Eq. (12) only as an educated guess and we have to prove that Eq. (14) is a noncontextuality inequality by showing that it holds whenever a P&M noncontextual model exists.

In order to construct an explicit violation of the noncontextuality inequality we can consider the equivalent of the standard quantum violation of the CHSH inequality: let $\mathcal{H}$ be a qubit Hilbert space, dim($\mathcal{H}$) = 2, let $\sigma_x, \sigma_y, \sigma_z$ be the Pauli matrices and let

$$
A_1 = \frac{1}{\sqrt{2}}(\sigma_x + \sigma_z), \quad A_2 = \frac{1}{\sqrt{2}}(\sigma_x - \sigma_z),
$$

$$
\sigma_{1+} = \frac{1}{2}|+\rangle\langle+|, \quad \sigma_{1-} = \frac{1}{2}|\rangle\langle-|,
$$

$$
\sigma_{2+} = \frac{1}{2}|0\rangle\langle0|, \quad \sigma_{2-} = \frac{1}{2}|1\rangle\langle1|,
$$

where $|+, \rangle$, $|-, \rangle$, and $|0, \rangle, |1, \rangle$ are the eigenbasis of the Pauli operators $\sigma_x, \sigma_z$ respectively. We have $\sigma_{1+} + \sigma_{1-} = \sigma_{2+} + \sigma_{2-}$, and so the constraint (13) is satisfied. We get

$$
\text{Tr}\{(A_1 + A_2)(\sigma_{1+} - \sigma_{1-})\} + \text{Tr}\{(A_1 - A_2)(\sigma_{2+} - \sigma_{2-})\} = 2\sqrt{2},
$$

hence the inequality (14) is violated.

Let us note that it was shown in Ref. [10] that stabilizer rebit theory, whose state space consists of the convex combinations of the states $|0, \rangle, |1, \rangle, |+, \rangle, |-, \rangle$ has a P&M noncontextual model which, at first sight, seems to contradict the presented violation of the noncontextuality inequality (14). There is no contradiction, however, because the observables $A_1$ and $A_2$ are not included in the stabilizer rebit theory.

**Conclusions.**— Our main results, Theorems 1 and 3, prove that contextuality is a precondition for entanglement and that only entangled states allow for robust remote preparations of contextuality. We have used these results to map noncontextuality inequalities to a class of entanglement witnesses in Proposition 4 and to obtain a class of noncontextuality inequalities from entanglement witnesses in Proposition 5.

Our results provide an insight into why entanglement is a precondition for secure quantum key distribution [3], since if there exists a preparation and measurement noncontextual hidden variable model, then the transmission channel between Alice and Bob can in principle be replaced by a classical channel by the eavesdropper, which makes secure key distribution impossible. As a consequence of our results any experiment which verifies entanglement of a state (e.g., by observing quantum steering or violation of a Bell inequality) immediately verifies P&M contextuality of the induced system. Moreover our results open the path to further transport of results between entanglement and contextuality, it is for example possible to take a contextuality-enabled task and transform it into a remote entanglement-enabled task. Thus our results provide a blueprint for connecting the resource theory of entanglement and contextuality.

We acknowledge support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, project numbers 447948357 and 440958198), the Sino-German Center for Research Promotion (Project...
Appendix A: Proof of Theorem 1

**Theorem 1.** Let $\rho_{AB}$ be a bipartite quantum state and assume that there exists a P&M noncontextual model for the set of states $\Lambda_A(\rho_{AB})$ and all possible measurements. Then, $\rho_{AB}$ is separable.

**Proof.** Let $\sigma_A \in \Lambda_A(\rho_{AB})$ and let $M = \{M_\lambda\}$ be a POVM on $\mathcal{H}_A$. According to our assumptions there exists P&M noncontextual hidden variable model for $\Lambda_A(\rho_{AB})$, thus we have $\text{Tr}(\sigma_A M_\lambda) = \sum_\lambda \text{Tr}(\sigma_A N_\lambda) \text{Tr}(\omega_\lambda M_\lambda)$ for some $\omega_\lambda \in \mathcal{D}(\mathcal{H}_A)$ and $N_\lambda \in \mathcal{B}(\mathcal{H}_A)$ such that $\text{Tr}(\sigma'_A N_\lambda) \geq 0$ for all $\sigma'_A \in \Lambda_A(\rho_{AB})$. Since $\sigma_A \in \Lambda_A(\rho_{AB})$ then there is some $E_B \geq 0$ such that $\sigma_A = \text{Tr}_B[(I_A \otimes E_B)\rho_{AB}]$ and we have $\text{Tr}(\sigma_A M_\lambda) = \text{Tr}[(M_\lambda \otimes E_B)\rho_{AB}]$ and $\text{Tr}(\sigma_A N_\lambda) = \text{Tr}[(N_\lambda \otimes E_B)\rho_{AB}]$. Moreover we will use that $\text{Tr}[(N_\lambda \otimes E_B)\rho_{AB}] = \text{Tr}(E_B \text{Tr}_A[(N_\lambda \otimes I_B)\rho_{AB}])$. Putting everything together we get

$$\text{Tr}[(M_\lambda \otimes E_B)\rho_{AB}] = \sum_\lambda \text{Tr}[(N_\lambda \otimes E_B)\rho_{AB}] \text{Tr}(\omega_\lambda M_\lambda)$$

$$= \sum_\lambda \text{Tr}[(M_\lambda \otimes N_\lambda \otimes E_B)(\omega_\lambda \otimes \rho_{AB})]$$

$$= \sum_\lambda \text{Tr}[(M_\lambda \otimes E_B)(\omega_\lambda \otimes \text{Tr}_A[(N_\lambda \otimes I_B)\rho_{AB}])].$$

Since this holds for every $M_\lambda$ and $E_B$ we must have $\rho_{AB} = \sum_\lambda \omega_\lambda \otimes \text{Tr}_A[(N_\lambda \otimes I_B)\rho_{AB}]$. In order to prove that $\rho_{AB}$ is separable we only need to prove that $\text{Tr}_A[(N_\lambda \otimes I_B)\rho_{AB}]$ is positive semidefinite for all $\lambda$, but this is straightforward as for any positive semidefinite $F_B \in \mathcal{B}(\mathcal{H}_B)$ we have

$$\text{Tr}\{F_B \text{Tr}_A[(N_\lambda \otimes I_B)\rho_{AB}]) = \text{Tr}\{N_\lambda \text{Tr}_B[(I_A \otimes F_B)\rho_{AB}]) = \mu \text{Tr}(N_\lambda \sigma'_A) \geq 0$$

where $\sigma'_A \in \Lambda_A(\rho_{AB})$ and $\mu \geq 0$ are such that $\text{Tr}_B[(I_A \otimes F_B)\rho_{AB}] = \mu \sigma'_A$. 



Appendix B: Proof of Theorem 2

Before proceeding to the proof of Theorem 2 we need to introduce two superoperators. Let $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ be a bipartite state, then the superoperator $\Psi^{\rho}_{A \rightarrow B} : \mathcal{B}(\mathcal{H}_A) \rightarrow \mathcal{B}(\mathcal{H}_B)$ is defined as follows: let $X_A \in \mathcal{B}(\mathcal{H}_A)$, then

$$\Psi^{\rho}_{A \rightarrow B}(X_A) = \text{Tr}_A[(X_A \otimes I_B)\rho_{AB}].$$

(18)

The superoperator $\Psi^{\rho}_{B \rightarrow A} : \mathcal{B}(\mathcal{H}_B) \rightarrow \mathcal{B}(\mathcal{H}_A)$ is defined analogously. Moreover we can express the remotely preparable sets using the superoperator $\Psi^{\rho}_{B \rightarrow A}$ as $\Lambda_A(\rho_{AB}) = \{\sigma_A \in \mathcal{D}(\mathcal{H}_A) : \sigma_A = \Psi^{\rho}_{B \rightarrow A}(E_B), E_B \geq 0\}$.

Let $\{X_i^A\}_{i=1}^{\dim(\mathcal{H}_A)^2}$ and $\{Y_j^B\}_{j=1}^{\dim(\mathcal{H}_B)^2}$ be an orthonormal operator basis of $\mathcal{B}(\mathcal{H}_A)$ and $\mathcal{B}(\mathcal{H}_B)$ respectively, i.e.,

$$\text{Tr}(X_i^A X_j^A) = \text{Tr}(Y_i^B Y_j^B) = \delta_{ij},$$

then the basis can be chosen such that $\rho_{AB} = \sum_{i=1}^{N} p_i X_i^A \otimes Y_i^B$, where $N = \max[\dim(\mathcal{H}_A)^2, \dim(\mathcal{H}_B)^2]$ and $p_i \geq 0$, this is essentially Schmidt decomposition applied to $\rho_{AB}$ as a vector in $\mathcal{B}(\mathcal{H}_A) \otimes \mathcal{B}(\mathcal{H}_B)$. Let $I_+ = \{i : p_i > 0\}$. We then define the superoperator $\Pi^{\rho}_{A} : \mathcal{B}(\mathcal{H}_A) \rightarrow \mathcal{B}(\mathcal{H}_A)$ as follows:

$$\Pi^{\rho}_{A} = \sum_{i=1}^{\dim(\mathcal{H}_A)^2} \alpha_i X_i^A = \sum_{i \in I_+} \alpha_i X_i^A.$$

(19)

$\Pi^{\rho}_{B} : \mathcal{B}(\mathcal{H}_B) \rightarrow \mathcal{B}(\mathcal{H}_B)$ is defined analogically.

The superoperators $\Pi^{\rho}_{A}$ and $\Pi^{\rho}_{B}$ will appear in our calculations for the following reason: the superoperator $\Psi^{\rho}_{B \rightarrow A}$ is not necessarily invertible, but we can invert it on its support. $\Pi^{\rho}_{A}$ and $\Pi^{\rho}_{B}$ are projections on the support of $\Psi^{\rho}_{B \rightarrow A}$ and so they will appear because we will use the pseudo-inverse of $\Psi^{\rho}_{B \rightarrow A}$.

**Lemma.** Let $\rho_{AB}$ be a separable bipartite quantum state with the decomposition $\rho_{AB} = \sum_\lambda \omega_\lambda \otimes K_\lambda$, where $\omega_\lambda \geq 0$, $\text{Tr}(\omega_\lambda) = 1$ and $K_\lambda \geq 0$. Assume that we have $\Pi^{\rho}_{B}(K_\lambda) \geq 0$ for all $\lambda$. Then there exists P&M noncontextual model for $\Lambda_A(\rho_{AB})$. 

M-0294), and from the ERC (Consolidator Grant 683107/TempoQ). MP is thankful for the financial support from the Alexander von Humboldt Foundation.
Proof. Let \( \{X_i^A\}_{i=1}^{\dim(H_A)^2} \) and \( \{Y_j^B\}_{j=1}^{\dim(H_B)^2} \) be orthonormal operator basis of \( B(H_A) \) and \( B(H_B) \) respectively such that \( \rho_{AB} = \sum_{i=1}^{N} p_i X_i^A \otimes Y_i^B \), where \( N = \max[\dim(H_A)^2, \dim(H_B)^2] \) and \( p_i \geq 0 \). Let \( \sigma_A \in \Lambda_A(\rho_{AB}) \) and let \( E_B \in B(H_B), \ E_B \geq 0 \), be the corresponding operator such that \( \sigma_A = Tr_B(\mathbf{1}_A \otimes E_B) \rho_{AB} \). Then \( E_B = \sum_{i=1}^{\dim(H_B)^2} \beta_i Y_i^B \) and we have
\[
\sigma_A = Tr_B([\mathbf{1}_A \otimes E_B)] \rho_{AB}) = \Psi_{B \rightarrow A}^\rho(E_B) = \sum_{i \in I_+} \beta_i p_i X_i^A.
\] (20)

We will now define a superoperator \( \Phi : B(H_A) \rightarrow B(H_B) \) that will act as the pseudo-inverse to the superoperator \( \Psi_{B \rightarrow A}^\rho \) as follows: let \( F_A \in B(H_A) \), then \( F_A = \sum_{i=1}^{\dim(H_A)^2} \alpha_i X_i^A \) and we define \( \Phi(F_A) = \sum_{i \in I_+} \alpha_i Y_i^B \). We clearly have
\[
(\Phi \circ \Psi_{B \rightarrow A}^\rho)(E_B) = \Pi_{B}^\rho(E_B),
\] (21)
where \( \circ \) denotes the concatenation of superoperators. Moreover it follows that \( \Psi_{B \rightarrow A}^\rho(E_B) = (\Psi_{B \rightarrow A}^\rho \circ \Pi_{B}^\rho)(E_B) \). Let \( M = \{M_i\} \) be a POVM, then we have
\[
Tr(\sigma_A M_a) = Tr([M_a \otimes E_B] \rho_{AB}) = Tr\{(M_a \otimes \Pi_{B}^\rho(E_B)) \rho_{AB}\} = \sum_{\lambda} Tr(\omega_{\lambda} M_a) Tr(\Pi_{B}^\rho(E_B) K_\lambda).
\] (22)

Using Eq. (21) we get
\[
Tr(\sigma_A M_a) = \sum_{\lambda} Tr[\Psi_{B \rightarrow A}^\rho(E_B) \Phi^*(K_\lambda)] Tr(\omega_{\lambda} M_a) = \sum_{\lambda} Tr[\sigma_A \Phi^*(K_\lambda)] Tr(\omega_{\lambda} M_a)
\] (23)
where \( \Phi^* \) is the adjoint superoperator to \( \Phi \). Denoting \( \Phi^*(K_\lambda) = N_\lambda \) we get
\[
Tr(\sigma_A M_a) = \sum_{\lambda} Tr(\sigma_A N_\lambda) Tr(\omega_{\lambda} M_a)
\] (24)
which is the desired result, we only need to check that \( N_\lambda \) satisfies the positivity and normalization conditions (3).

Let \( \sigma_A \in \Lambda_A(\rho_{AB}) \) be given as \( \sigma_A = \Psi_{B \rightarrow A}^\rho(E_B) \) for some \( E_B \in B(H_B), \ E_B \geq 0 \), then we have
\[
Tr(\sigma_A N_\lambda) = Tr[\Phi(\sigma_A) K_\lambda] = Tr[\Pi_{B}^\rho(E_B) K_\lambda] = Tr[E_B \Pi_{B}^\rho(K_\lambda)] \geq 0
\] (25)
since \( \Pi_{B}^\rho(K_\lambda) \geq 0 \) according to the assumptions of the theorem. To check normalization note that due to the adjoint version of Eq. (21) we have that \( \Phi^* \) is the pseudo-inverse of the superoperator \( \Psi_{A \rightarrow B}^\rho \). Using \( \sum_{\lambda} K_\lambda = Tr_A(\rho_{AB}) = Tr_A([\mathbf{1}_A \otimes I_B] \rho_{AB}) \) we get that \( \Phi^*(\sum_{\lambda} K_\lambda) = \Pi_{A}^\rho(\mathbf{1}_A) \) and so we have \( \sum_{\lambda} N_\lambda = \Phi^*(\sum_{\lambda} K_\lambda) = \Pi_{A}^\rho(\mathbf{1}_A) \). It follows that \( \sum_{\lambda} Tr(\sigma_A N_\lambda) = Tr[\sigma_A \Pi_{A}^\rho(\mathbf{1}_A)] = Tr(\sigma_A) = 1 \) because \( \Pi_{A}^\rho(\sigma_A) = \sigma_A \). 

Theorem 2. Let \( \rho_{AB} \) be a separable bipartite quantum state with the decomposition \( \rho_{AB} = \sum_{\lambda} \omega_\lambda \otimes K_\lambda \), where \( \omega_\lambda \geq 0 \), \( Tr(\omega_\lambda) = 1 \) and \( K_\lambda \geq 0 \). Assume that \( K_\lambda \) belongs to the linear hull of \( \Lambda_B(\rho_{AB}) \) for all \( \lambda \). Then there exists PE\( \mathcal{M} \) noncontextual model for \( \Lambda_B(\rho_{AB}) \).

Proof. The result follows from the previous lemma. Since for every \( \sigma \in \Lambda_B(\rho_{AB}) \) we have \( \Pi_{B}^\rho(\sigma) = \sigma \), it follows that if \( K_\lambda \) belongs to the linear hull of \( \Lambda_B(\rho_{AB}) \) then we have \( \Pi_{B}^\rho(K_\lambda) = K_\lambda \geq 0 \).

Appendix C: Proof of Theorem 3

Theorem 3. Let \( \rho_{AB} \) be a separable quantum state. Then for almost all separable quantum states \( \tau_{AB} \) there is a \( \delta(\tau_{AB}) > 0 \) depending on \( \tau_{AB} \), such that for every \( \varepsilon \in (0, \delta) \) there exists PE\( \mathcal{M} \) noncontextual model for \( \Lambda_{A}[1 - \varepsilon \rho_{AB} + \varepsilon \tau_{AB}] \).

Proof. We will assume that \( n = \dim(H_A) = \dim(H_B) \) as we can always embed smaller Hilbert spaces into larger ones. And let \( \tau_{AB} \in D(H_A \otimes H_B) \) be a randomly selected separable state. Since \( \dim(H_A) = \dim(H_B) = n \), we can represent the superoperators \( \Psi_{B \rightarrow A}^\rho \) and \( \Psi_{B \rightarrow A}^\tau \) by \( n^2 \times n^2 \) matrices \( M(\Psi_{B \rightarrow A}^\rho) \) and \( M(\Psi_{B \rightarrow A}^\tau) \). Moreover since \( \tau_{AB} \) is randomly selected, we have \( \det[ M(\Psi_{B \rightarrow A}^\tau) ] \neq 0 \), since almost all matrices have non-zero determinants.
It follows that the matrix corresponding to the superoperator \( \Psi^{(1-\varepsilon)\rho+\varepsilon\tau}_{B\rightarrow A} \) is \((1-\varepsilon)M(\Psi^0_{B\rightarrow A}) + \varepsilon M(\Psi^1_{B\rightarrow A})\). In order to finish the proof we want to show that there is some \( \delta \in (0,1) \) such that for all \( \varepsilon \in (0,\delta) \) we have \( \det[(1-\varepsilon)M(\Psi^0_{B\rightarrow A}) + \varepsilon M(\Psi^1_{B\rightarrow A})]\) \(\neq 0\), then it follows that the superoperator \( \Psi^{(1-\varepsilon)\rho+\varepsilon\tau}_{B\rightarrow A} \) is invertible.

We have that \( \det[(1-\varepsilon)M(\Psi^0_{B\rightarrow A}) + \varepsilon M(\Psi^1_{B\rightarrow A})] \) is either constant in \( \varepsilon \) or a polynomial of finite order. Since the superoperator \( \Psi^0_{B\rightarrow A} \) is invertible we must have \( \det[M(\Psi^1_{B\rightarrow A})] \neq 0 \). Therefore if \( \det[(1-\varepsilon)M(\Psi^0_{B\rightarrow A}) + \varepsilon M(\Psi^1_{B\rightarrow A})] \) is constant, then \( \det[(1-\varepsilon)M(\Psi^0_{B\rightarrow A}) + \varepsilon M(\Psi^1_{B\rightarrow A})] \neq 0 \) for all \( \varepsilon \in [0,1] \) and the result follows. If \( \det[(1-\varepsilon)M(\Psi^0_{B\rightarrow A}) + \varepsilon M(\Psi^1_{B\rightarrow A})] \) is not constant, then let \( \delta \) be the smallest root of \( \det[(1-\varepsilon)M(\Psi^0_{B\rightarrow A}) + \varepsilon M(\Psi^1_{B\rightarrow A})] \) from the interval \((0,1)\). It follows that for all \( \varepsilon \in (0,\delta) \) we have \( \det[(1-\varepsilon)M(\Psi^0_{B\rightarrow A}) + \varepsilon M(\Psi^1_{B\rightarrow A})] \neq 0 \) and thus the corresponding map is invertible.

\[ \square \]

One can construct suitable \( \tau_{AB} \) explicitly as follows: Let \( \frac{1}{\sqrt{\varepsilon_i}} X_i^A \in \mathcal{B}(\mathcal{H}_A) \) and \( \frac{1}{\sqrt{\varepsilon_i}} Y_i^B \in \mathcal{B}(\mathcal{H}_B) \) be orthonormal bases, i.e., \( \text{Tr}(X_i^A) = \text{Tr}(Y_i^B) = 0 \) and \( \text{Tr}(X_i^A X_j^A) = \text{Tr}(Y_i^B Y_j^B) = \delta_{ij} \). Note that these are different basis than the Schmidt basis used in previous constructions. We then define \( \tau_{AB} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) \) as \( \tau_{AB} = \frac{1}{\sqrt{\varepsilon_i}} X_i^A \otimes Y_i^B \), where \( \mu > 0 \). It follows that we can always choose \( \mu \) so small that \( \tau_{AB} \geq 1 \), moreover such that \( \tau_{AB} \) is a separable state. Let \( E_B \in \mathcal{B}(\mathcal{H}_B) \), then \( E_B = \beta \frac{1}{\sqrt{n}} + \sum_i \beta_i Y_i^B \) and we have \( \text{Tr}_B((I_A \otimes E_B) \tau_{AB}) = \beta_1 \frac{1}{\sqrt{n}} + \sum_i \beta_i X_i^A \), from which it follows that the superoperator \( \Psi^\tau_{B\rightarrow A} \) is invertible. It then follows that \( \Psi^{(1-\varepsilon)\rho+\varepsilon\tau}_{B\rightarrow A} \) is also invertible for \( \varepsilon \in (0,\delta) \) for suitable choice of \( \delta \).

\[ \square \]

**Appendix D: Proof of Proposition 4**

The following result is a restatement of a known noncontextuality inequality [25], our only modification is that we allow for unnormalized states.

**Proposition.** Let \( K \subset \mathcal{D}(\mathcal{H}) \) be the set of allowed preparations and let \( \text{cone}(K) \) denote the set of all unnormalized allowed preparations, that is all operators of the form \( \mu \hat{\sigma} \), where \( \mu \in \mathbb{R}, \mu \geq 0 \) and \( \hat{\sigma} \in K \). Let \( \sigma_{t,b} \in \text{cone}(K) \) for \( t \in \{1,2,3\} \) and \( b \in \{0,1\} \) be such that \( \sigma_* = \frac{1}{2} (\sigma_{t,0} + \sigma_{t,1}) = \frac{1}{2} (\sigma_{t',0} + \sigma_{t',1}) \) for all \( t,t' \in \{1,2,3\} \) and \( M_{t,b} \in \mathcal{B}(\mathcal{H}) \) be positive operators, \( M_{t,b} \geq 0 \), such that \( \frac{1}{3} \sum_{t=1}^3 M_{t,b} = \frac{1}{2} \) and \( M_{t,0} + M_{t,1} = I \) for all \( t \). If there is preparation and measurement nocontextual hidden variable model for \( K \), then we have

\[ \frac{3}{2} \sum_{t=1}^3 \sum_{b=0}^1 \text{Tr}(\sigma_{t,b} M_{t,b}) \leq 5 \text{Tr}(\sigma_*) \tag{26} \]

**Proof.** Using Eq. (4) we get

\[
\sum_{t=1}^3 \sum_{b=0}^1 \text{Tr}(\sigma_{t,b} M_{t,b}) = \frac{3}{2} \sum_{t=1}^3 \sum_{b=0}^1 \text{Tr}(\sigma_{t,b} N_{t,b}) \text{Tr}(\omega_{t} M_{t,b}) \leq \frac{3}{2} \sum_{t=1}^3 \sum_{b=0}^1 \text{max}_t \text{Tr}(\omega_{t} M_{t,b}) \sum_{b=0}^1 \text{Tr}(\sigma_{t,b} N_{t,b})
\]

\[ = 2 \sum_{t=1}^3 \text{max}_b \text{Tr}(\omega_{t} M_{t,b}) \sum_{b=0}^1 \text{Tr}(\sigma_{t,b} N_{t,b}) \leq 6 \text{max} \left( \frac{1}{2} \sum_{t=1}^3 \text{max}_b \text{Tr}(\omega_{t} M_{t,b}) \right) \sum_{t=1}^3 \text{Tr}(\sigma_*) \]

\[ \leq 5 \text{Tr}(\sigma_*) \tag{27} \]

where we have used that \( \frac{1}{2} \sum_{t=1}^3 \text{max}_b \text{Tr}(\omega_{t} M_{t,b}) \leq \frac{5}{6} \), which is a standard step in proving the origin noncontextuality inequality, see [9, 25]. To see that \( \sum_{t=1}^3 \text{Tr}(\sigma_N) = \text{Tr}(\sigma_*) \) simply note that \( \sigma_* \in \text{cone}(K) \) and so \( \sigma_* = \text{Tr}(\sigma_*) \hat{\sigma}_* \) where \( \hat{\sigma}_* \in K \). Then \( \sum_{t=1}^3 \text{Tr}(\sigma_{t,b} N_{t,b}) = \text{Tr}(\sigma_*) \sum_{t=1}^3 \text{Tr}(\hat{\sigma}_* N_{t,b}) = \text{Tr}(\sigma_*) \) as a result of the normalization of \( N \).

**Proposition 4.** Let \( \rho_{AB} \) be a separable quantum state. Let \( E_{t,b} \) and \( M_{t,b} \) be positive operators, \( M_{t,b} \geq 0 \) and \( M_{t,b} \geq 0 \), such that \( E_* = \frac{1}{2} (E_{t,0} + E_{t,1}) \), \( \frac{1}{3} \sum_{t=1}^3 M_{t,b} = \frac{1}{2} \) and \( I_A = M_{t,0} + M_{t,1} \) for all \( t \in \{1,2,3\} \) and \( b \in \{0,1\} \). Then

\[ \sum_{t=1}^3 \sum_{b=0}^1 \text{Tr}[(M_{t,b} \otimes E_{t,b}) \rho_{AB}] \leq 5 \text{Tr}[(I_A \otimes E_*) \rho_{AB}] \tag{7} \]

Moreover, there is an entangled state \( \rho_{AB} \) that violates Eq. (7) for suitable choice of the operators \( E_{t,b} \) and \( M_{t,b} \).
Proof. The proof is straightforward: assume that \( \rho_{AB} \) is separable and let \( \tau_{AB} \) be a separable state such that there exists preparation and measurement noncontextual hidden variable model for \( \Lambda N(1-\varepsilon)\rho_{AB}+\varepsilon\tau_{AB} \) for all \( \varepsilon \in (0, \delta) \) for suitable \( \delta \in (0, 1) \). Denote \( \sigma_{t,b} = \text{Tr}_t((1_A \otimes \mathbb{E}_b)((1-\varepsilon)\rho_{AB} + \varepsilon\tau_{AB})) \) and \( \sigma_{*,*} = \text{Tr}_B((1_A \otimes \mathbb{E}_*)((1-\varepsilon)\rho_{AB} + \varepsilon\tau_{AB})) \), then Eq. (6) becomes \( \sum_{t=1}^{3}\sum_{b=0}^{1}\text{Tr}(\sigma_{t,b} \mathbb{M}_{t,b}) \leq 5\text{Tr}(\sigma_{*,*}) \). We thus get

\[
\sum_{t=1}^{3}\sum_{b=0}^{1}\text{Tr}([E_{t,b} \otimes M_{t,b}] (1-\varepsilon)\rho_{AB} + \varepsilon\tau_{AB})) \leq 5\text{Tr}([E_* \otimes 1_B] (1-\varepsilon)\rho_{AB} + \varepsilon\tau_{AB})) \tag{28}
\]

for all \( \varepsilon \in (0, \delta) \). Taking the limit \( \varepsilon \to 0^+ \) yields Eq. (7).

To show that there is an entangled state that violates Eq. (7) simply assume that \( \sigma_{t,b} \) and \( M_{t,b} \) are states and POVMs that violate Eq. (7), it is known that such states and POVMs exist \([9, 25]\). Let \( |\phi^+\rangle = \frac{1}{\sqrt{\dim(\mathcal{H})}} \sum_{i=1}^{\dim(\mathcal{H})} |ii\rangle \) be the maximally entangled state and take \( \rho_{AB} = |\phi^+\rangle\langle \phi^+ | \). Take \( E_{t,b} = \dim(\mathcal{H})\sigma_{t,b}^T \), where \( A^T \) is the transposition of \( A \) with respect to the basis \( |i\rangle \). Then we have \( \text{Tr}_B(\mathbb{1}_A \otimes E_{t,b})\rho_{AB} = \sigma_{t,b} \) and so Eq. (7) must be violated because the corresponding noncontextuality inequality is violated. \( \square \)

Appendix E: Proof of Proposition 5

Proposition 5. Let \( K \) be a set of allowed preparations. Let \( \sigma_* \in K \) and let \( i \in \{1, 2\} \), let \( \sigma_{i+}, \sigma_{i-}, \sigma_{10} \in \text{cone}(K) \) be subnormalized preparations such that

\[
\sigma_{1+} + \sigma_{i-} + \sigma_{10} = \sigma_* = \sigma_{2+} + \sigma_{2-} + \sigma_{20}. \tag{13}
\]

Let \( A_i \) be observables such that \(-1 \leq A_i \leq 1 \) for all \( i \in \{1, 2\} \). If there is a P\&M contextual model for \( K \), then

\[
\text{Tr}([A_1 + A_2](\sigma_{1+} - \sigma_{1-})) + \text{Tr}([A_1 - A_2](\sigma_{2+} - \sigma_{2-})) \leq 2. \tag{14}
\]

Proof. In order to shorten the notation let us denote

\[
\Xi(A, \sigma) = \text{Tr}([A_1 + A_2](\sigma_{1+} - \sigma_{1-})) + \text{Tr}([A_1 - A_2](\sigma_{2+} - \sigma_{2-})). \tag{29}
\]

The proof is straightforward: assume that there is preparation and measurement noncontextual hidden variable model for \( K \subset D(\mathcal{H}) \). Then due to the linearity of Eq. (4) we have \( \text{Tr}(\sigma_{t,b} A_j) = \sum_\lambda \text{Tr}(\sigma_{t,b} \lambda) \omega_j(A_j) \sigma_{t,b} \) for all \( i, j \in \{1, 2\} \), this follows for example from \( A_i = 2M_i - 1 \) where \( 0 \leq M_i \leq 1 \) is an appropriate operator. We thus get

\[
\Xi(A, \sigma) = \sum_\lambda \text{Tr}[N(\sigma_{1+} - \sigma_{1-})]\text{Tr}([A_1 + A_2]\omega_\lambda] + \sum_\lambda \text{Tr}[N(\sigma_{2+} - \sigma_{2-})]\text{Tr}([A_1 - A_2]\omega_\lambda] = \sum_\lambda \text{Tr}[N(\sigma_{1+} - \sigma_{1-} + \sigma_{2+} - \sigma_{2-})]\text{Tr}([A_1\omega_\lambda] + \sum_\lambda \text{Tr}[N(\sigma_{1+} - \sigma_{1-} - \sigma_{2+} + \sigma_{2-})]\text{Tr}([A_2\omega_\lambda] \leq \sum_\lambda \text{Tr}[N(\sigma_{1+} - \sigma_{1-} + \sigma_{2+} - \sigma_{2-})] + \sum_\lambda \text{Tr}[N(\sigma_{1+} - \sigma_{1-} - \sigma_{2+} + \sigma_{2-})], \tag{30}
\]

where we have used that \( \text{Tr}(A_i \omega_\lambda) \in [-1, 1] \) for all \( i \in \{1, 2\} \) and all \( \lambda \). Let us fix \( \lambda \) and let us inspect the term \( \text{Tr}[N(\sigma_{1+} - \sigma_{1-} + \sigma_{2+} - \sigma_{2-})] + \text{Tr}[N(\sigma_{1+} - \sigma_{1-} - \sigma_{2+} + \sigma_{2-})] = \xi_\lambda \). There are four options on how the signs can be assigned: if \( \text{Tr}[N(\sigma_{1+} - \sigma_{1-} + \sigma_{2+} - \sigma_{2-})] \geq 0 \) and \( \text{Tr}[N(\sigma_{1+} - \sigma_{1-} - \sigma_{2+} + \sigma_{2-})] \geq 0 \) we get

\[
\xi_\lambda = \text{Tr}[N(\sigma_{1+} - \sigma_{1-} + \sigma_{2+} - \sigma_{2-})] + \text{Tr}[N(\sigma_{1+} - \sigma_{1-} - \sigma_{2+} + \sigma_{2-})] = 2\text{Tr}[N(\sigma_{1+} - \sigma_{1-})] \leq 2\text{Tr}(N(\sigma_*)), \tag{31}
\]

where the last inequality follows from Eq. (13) and from the positivity condition in Eq. (3). If \( \text{Tr}[N(\sigma_{1+} - \sigma_{1-} + \sigma_{2+} - \sigma_{2-})] \geq 0 \) and \( \text{Tr}[N(\sigma_{1+} - \sigma_{1-} - \sigma_{2+} + \sigma_{2-})] < 0 \) we get

\[
\xi_\lambda = \text{Tr}[N(\sigma_{1+} - \sigma_{1-} + \sigma_{2+} - \sigma_{2-})] - \text{Tr}[N(\sigma_{1+} - \sigma_{1-} - \sigma_{2+} + \sigma_{2-})] = 2\text{Tr}[N(\sigma_{2+} - \sigma_{2-})] \leq 2\text{Tr}(N(\sigma_*)), \tag{32}
\]

if \( \text{Tr}[N(\sigma_{1+} - \sigma_{1-} + \sigma_{2+} - \sigma_{2-})] < 0 \) and \( \text{Tr}[N(\sigma_{1+} - \sigma_{1-} - \sigma_{2+} + \sigma_{2-})] \geq 0 \) we get

\[
\xi_\lambda = -\text{Tr}[N(\sigma_{1+} - \sigma_{1-} + \sigma_{2+} - \sigma_{2-})] + \text{Tr}[N(\sigma_{1+} - \sigma_{1-} - \sigma_{2+} + \sigma_{2-})] = 2\text{Tr}[N(\sigma_{2-} - \sigma_{2+})] \leq 2\text{Tr}(N(\sigma_*)), \tag{33}
\]

if \( \text{Tr}[N(\sigma_{1+} - \sigma_{1-} + \sigma_{2+} - \sigma_{2-})] < 0 \) and \( \text{Tr}[N(\sigma_{1+} - \sigma_{1-} - \sigma_{2+} + \sigma_{2-})] < 0 \) we get

\[
\xi_\lambda = -\text{Tr}[N(\sigma_{1+} - \sigma_{1-} + \sigma_{2+} - \sigma_{2-})] - \text{Tr}[N(\sigma_{1+} - \sigma_{1-} - \sigma_{2+} + \sigma_{2-})] = 2\text{Tr}[N(\sigma_{1+} - \sigma_{1-})] \leq 2\text{Tr}(N(\sigma_*)). \tag{34}
\]

We thus have \( \Xi(A, \sigma) \leq 2\sum_\lambda \text{Tr}(N(\sigma_*)) = 2 \) where we have used that \( \sigma_* \in K \) and the normalization condition in Eq. (3). \( \square \)
[1] O. Gühne and G. Tóth, Physics Reports 474, 1 (2009), arXiv:0811.2803.
[2] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Reviews of Modern Physics 81, 865 (2009), arXiv:quant-ph/0702225.
[3] M. Curty, M. Lewenstein, and N. Lu, Physical Review Letters 92, 217903 (2004), arXiv:quant-ph/0307151.
[4] L. Pezzè and A. Smerzi, Physical Review Letters 102, 10.1103/PhysRevLett.102.100401 (2007), arXiv:0711.4840.
[5] G. Tóth and I. Apellaniz, Journal of Physics A: Mathematical and Theoretical 47, 424006 (2014), arXiv:1405.4878.
[6] R. Uola, A. C. S. Costa, H. C. Nguyen, and O. Gühne, Reviews of Modern Physics 92, 015001 (2020), arXiv:1903.06663.
[7] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Reviews of Modern Physics 86, 419 (2014), arXiv:1303.2849.
[8] R. W. Spekkens, Physical Review A 71, 052108 (2005), arXiv:quant-ph/0406166.
[9] C. Budroni, A. Cabello, O. Gühne, M. Kleinmann, and J.-A. Larsson, arXiv:2102.13036 (2021).
[10] D. Schmid, J. H. Selby, E. Wolfe, R. Kunjwal, and R. W. Spekkens, PRX Quantum 2, 010331 (2021), arXiv:1911.10386.
[11] D. Schmid, J. H. Selby, M. F. Pusey, and R. W. Spekkens, arXiv:2005.07161 (2020).
[12] J. H. Selby, D. Schmid, E. Wolfe, A. B. Sainz, R. Kunjwal, and R. W. Spekkens, arXiv:2106.09045 (2021).
[13] J. H. Selby, D. Schmid, E. Wolfe, A. B. Sainz, R. Kunjwal, and R. W. Spekkens, arXiv:2112.04521 (2021).
[14] D. Schmid and R. W. Spekkens, Physical Review X 8, 011015 (2018), arXiv:1706.04588.
[15] R. W. Spekkens, D. H. Buzacott, A. J. Keelh, B. Toner, and G. J. Pryde, Physical Review Letters 102, 010401 (2009), arXiv:0805.1463.
[16] C. H. Bennett, G. Brassard, and N. D. Mermin, Physical Review Letters 68, 557 (1992).
[17] M. Plávala, Journal of Physics A: Mathematical and Theoretical 55, 174001 (2021), arXiv:2112.10596.
[18] A. Tavakoli and R. Uola, Physical Review Research 2, 013011 (2020), arXiv:1905.03614.
[19] C. H. Bennett, D. P. DiVincenzo, P. W. Shor, J. A. Smolin, B. M. Terhal, and W. K. Wootters, Physical Review Letters 87, 077902 (2001), arXiv:quant-ph/0006044.
[20] S. Jevtic, M. Pusey, D. Jennings, and T. Rudolph, Physical Review Letters 113, 020402 (2014), arXiv:1303.4724.
[21] V. Gitton and M. P. Woods, Quantum 6, 732 (2022), arXiv:2003.06426.
[22] T. Heinosaari and M. Ziman, The Mathematical Language of Quantum Theory. From Uncertainty to Entanglement (Cambridge University Press, 2012).
[23] H. C. Nguyen, H.-V. Nguyen, and O. Gühne, Physical Review Letters 122, 240401 (2019), arXiv:1808.09349.
[24] R. T. Rockafellar, Convex Analysis, Princeton Landmarks in Mathematics and Physics (Princeton University Press, 1997).
[25] M. D. Mazurek, M. F. Pusey, R. Kunjwal, K. J. Resch, and R. W. Spekkens, Nature Communications 7, ncomms11780 (2016), arXiv:1505.06244.
[26] G. Tóth, Physical Review A 71, 010301 (2005), arXiv:quant-ph/0406061.
[27] C. Brukner and V. Vedral, arXiv:quant-ph/0406040 (2004).
[28] M. R. Dowling, A. C. Doherty, and S. D. Bartlett, Physical Review A 70, 062113 (2004), arXiv:quant-ph/0408086.
[29] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Physical Review Letters 23, 880 (1969).
[30] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Physical Review Letters 24, 549 (1970).