

# Noise-robust preparation contextuality shared between any number of observers via unsharp measurements

Hammad Anwer<sup>1</sup>, Natalie Wilson<sup>1</sup>, Ralph Silva<sup>2</sup>, Sadiq Muhammad<sup>1</sup>, Armin Tavakoli<sup>3,4,5</sup>, and Mohamed Bourenane<sup>1</sup>

<sup>1</sup>Department of Physics, Stockholm University, S-10691 Stockholm, Sweden

<sup>2</sup>Institute for Theoretical Physics, ETH Zurich, Switzerland

<sup>3</sup>Département de Physique Appliquée, Université de Genève, CH-1211 Genève, Switzerland

<sup>4</sup>Institute for Quantum Optics and Quantum Information - IQOQI Vienna, Austrian Academy of Sciences, Boltzmannngasse 3, 1090 Vienna, Austria

<sup>5</sup>Institute for Atomic and Subatomic Physics, Vienna University of Technology, 1020 Vienna, Austria

**Multiple observers who independently harvest nonclassical correlations from a single physical system share the system's ability to enable quantum correlations. We show that any number of independent observers can share the preparation contextual outcome statistics enabled by state ensembles in quantum theory. Furthermore, we show that even in the presence of any amount of white noise, there exists quantum ensembles that enable such shared preparation contextuality. The findings are experimentally realised by applying sequential unsharp measurements to an optical qubit ensemble which reveals three shared demonstrations of preparation contextuality.**

## 1 Introduction

Quantum correlations can surpass the limitations of corresponding classical models. In their most common form, quantum correlations are obtained from the outcomes of *single* (albeit randomly chosen) measurements performed on a physical system. After the measurement, the physical system can be discarded, or even demolished by the measurement apparatus. Therefore, since one does not need to consider the measurement-induced decoherence in the state of the physical system, optimal quantum correlations are often obtained from sharp (projective) measurements that extract a maximal amount of information from the physical system while also inducing a maximal disturbance in its state [1].

Arguably, the fact that measurements disturb physical states should have interesting consequences for more general quantum correlations. To reveal the influence of measurement-induced disturbances on observed outcome statistics, one requires systems to undergo more than a single measurement. A simple scenario for studying the trade-off between the strength of quantum correlations and the disturbance induced

by extracting them is one in which quantum correlations are *shared* between many observers. Sharing quantum correlations means that a physical system is measured by a sequence of independent observers, each of whom are tasked with falsifying the existence of a classical model for their observed correlations. Hence, the stronger the correlations extracted by the first observer, the larger the disturbance induced in the state of the system, and thus the weaker the correlations that can possibly be extracted by a second observer. Sharing quantum correlations requires the first observer to measure in such a way that the outcome correlations are strong enough to elude all classical models while the induced disturbance is small enough to enable a second observer independently repeat the same feat. Understanding and characterising quantum correlations obtained via sequential measurements is a conceptually interesting problem [2–6] which has promising applications in quantum information protocols [7–9].

Sharing quantum correlations was first studied in the context of Bell inequality tests [4] where it was found that a pair of qubits in a singlet state can enable two sequential Bell inequality violations. This has also been experimentally demonstrated [10, 11]. Moreover, shared quantum correlations have recently also been studied in other tasks such as entanglement witnessing [12], quantum steering [13, 14] and a semi-device-independent setting [15–18].

Here, we theoretically and experimentally study the sharing of quantum correlations that demonstrate preparation contextuality. These are correlations that cannot be reproduced in a hidden variable theory that ascribes equivalent representations to indistinguishable preparations, i.e. it disregards the context (specific procedure) underlying a state preparation [19]. Such quantum contextuality does not require entanglement but only single quantum systems, and is well-studied both in theory (see e.g. Refs. [19–28]) and experiment (see e.g. Refs. [21, 29, 30]). In our scenario, states are sampled from an ensemble and commu-

nicated sequentially between independent observers, each of whom performs a measurement with the aim of obtaining preparation contextual outcome statistics. We show that preparation contextuality can be shared between any number of sequential observers. Furthermore, we show that the sharing is robust to noise, in the sense that for any given number of independent observers and exposure to any nontrivial amount of white noise, one can find an ensemble whose contextuality can be shared between all the observers. We proceed to experimentally demonstrate the sharing of preparation contextuality. We realise a four-observer scenario in which the first observer prepares an optical qubit ensemble and the remaining three observers perform sequential unsharp (non-maximally disturbing) measurements. Thus, we obtain three shared demonstrations of preparation contextuality.

## 2 Nonclassicality via preparation contextuality

The impossibility of describing the set of observables in quantum theory by underlying classical (noncontextual) quantities originates in the arguments of Bell, Kochen and Specker [31]. More recently, the notion of contextuality has seen a generalisation formulated in operational terms (i.e., in terms of probabilities) applying to measurements, transformations and preparations [19]. Here, we are interested in contextuality in terms of preparations.

The predictions of an operational theory (e.g. quantum theory) may be explained by an ontological model [32]. An ontological model ascribes a set  $\Lambda$  of ontic (objective) states  $\lambda$  to each physical system  $S$ . A particular preparation  $P$  of the system is associated to a distribution  $\mu_P(\lambda)$  over the ontic state space. Similarly, the probability of outcome  $b$  of a measurement  $M$  is described by a response function  $\xi_{b,M}(\lambda)$ . The ontological model thus seeks a  $\mu$  and a  $\xi$  to explain the observed statistics by  $p(b|P, M) = \int_{\Lambda} \mu_P(\lambda) \xi_{b,M}(\lambda) d\lambda$ . Note that such an ontological model only concerns the observed probabilities and does not need to explicitly reference an underlying physical theory. The model is said to be *preparation noncontextual* if two different preparations  $P$  and  $P'$  that cannot be distinguished by the statistics generated by any measurement (that is;  $\forall M : p(b|P, M) = p(b|P', M)$ ) are associated to the *same* distribution over ontic states, i.e.,  $\mu_P = \mu_{P'}$ . If observed statistics falsify this assumption, the it is said to be preparation *contextual*. Quantum state ensembles are known to enable preparation contextuality.

In order to prove preparation contextuality, it is sufficient to violate an inequality bounding the correlations attainable by any preparation noncontextual model. We focus on a family of such inequalities in-

troduced in Ref. [21] related to a variant of Random Access Coding [33, 34]. Consider a party Alice receiving a random input string  $x = x_1 \dots x_n \in \{0, 1\}^n$ . Her input is associated to a preparation  $P_x$  (one of  $2^n$  possible) which is sent to a receiver Bob. Her preparations are constrained to satisfy certain indistinguishability relations: there must exist no measurement that can reveal any information about the parity of the string  $r \cdot x$  for every  $r \in \{0, 1\}^n$  with  $|r| \geq 2$ . Bob receives a random input  $y \in \{1, \dots, n\}$ , and performs a measurement  $\{M_y^b\}$  with outcome  $b \in \{0, 1\}$ . The partnership is awarded a point if the outcome of Bob coincides with the  $y$ th entry in Alice's string. In any preparation noncontextual theory, the probability of winning obeys the following bound [21]:

$$\mathcal{A}^{(n)} \equiv \frac{1}{n2^n} \sum_{x,y} p(b = x_y | x, y) \leq \frac{n+1}{2n}. \quad (1)$$

Due to the contextual nature of quantum theory, these inequalities can be violated. Maximal quantum violations for any  $n \geq 2$  are known [35]. Bob performs dichotomic measurements characterised by an observable  $G_{n,y}^T$ , where  $T$  denotes transpose. These are recursively defined from  $G_{2,1} = \sigma_x$ ,  $G_{2,2} = \sigma_y$ , and  $G_{3,1} = \sigma_x$ ,  $G_{3,2} = \sigma_y$  and  $G_{3,3} = \sigma_z$ , and

$$\begin{aligned} n \text{ even: } & G_{n,k} = G_{n-1,k} \otimes \sigma_x \quad \forall k \in \{1, \dots, n-1\}, \\ n \text{ odd: } & G_{n,k} = G_{n-2,k} \otimes \sigma_x \quad \forall k \in \{1, \dots, n-2\} \end{aligned} \quad (2)$$

with  $G_{n,n} = \mathbb{1} \otimes \sigma_y$  if  $n > 3$  is even, and  $G_{n,n} = \mathbb{1} \otimes \sigma_z$  and  $G_{n,n-1} = \mathbb{1} \otimes \sigma_y$  if  $n > 3$  is odd. Note that the dimension of  $G_{n,k}$  is  $2^{\lfloor n/2 \rfloor}$ . The optimal preparations are states of  $\lfloor n/2 \rfloor$  qubits specified by

$$\rho_x = \text{tr}_A \left[ (\mathbb{1} + A_x) \otimes \mathbb{1} (|\phi_{\max}\rangle\langle\phi_{\max}|)^{\otimes \lfloor n/2 \rfloor} \right], \quad (3)$$

where  $A_x = \frac{1}{\sqrt{n}} \sum_{i=1}^n (-1)^{x_i} G_{n,i}$ ,  $|\phi_{\max}\rangle = (|0,0\rangle + |1,1\rangle)/\sqrt{2}$ , and the trace is taken over the first system in every entangled pair. Note that Alice's preparations are single quantum systems, and only for simplicity written in terms of post-measurement states of a collection of entangled states. The presented strategy leads to the maximal quantum value  $\mathcal{A}^{(n)} = 1/2(1 + 1/\sqrt{n})$  for every  $n$  [35].

## 3 Sequential scenario

We consider a scenario in which the ability to violate the inequality (1) is shared between many independent observers, named  $\text{Bob}_1, \dots, \text{Bob}_m$ , each of whom receive an independent random input  $y_k \in \{1, \dots, n\}$  and output  $b_k \in \{0, 1\}$ . Alice's randomly chosen preparation is sent to  $\text{Bob}_1$  who performs a measurement and passes the post-measurement state to  $\text{Bob}_2$  who performs a measurement and passes the post-measurement state to  $\text{Bob}_3$  etc. The scenario is

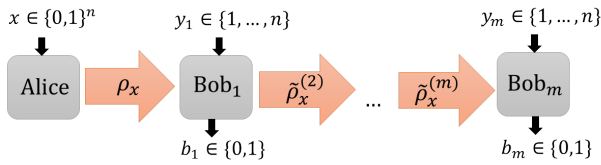


Figure 1: Alice's preparations are sent from one observer to the next, each performing a measurement aiming to independently reveal preparation contextual statistics. To this end, only the average post-measurement state  $\tilde{\rho}_x^{(k)}$  is relevant.

illustrated in Fig. 1. The pair Alice-Bob<sub>k</sub> uses the marginal distribution  $p(b_k|x, y_k)$  to compute the witness (1) (here labelled  $\mathcal{A}_k^{(n)}$ ) to check for preparation contextuality.

In a quantum approach, we may denote Alice's preparations by  $\rho_x$  which must satisfy the indistinguishability relation  $\sum_{r \cdot x=0} \rho_x = \sum_{r \cdot x=1} \rho_x$  for every string  $r$  with  $|r| \geq 2$ . Since one has to keep track of both the statistics and the post-measurement states of each Bob, we require the detailed set of Kraus operators for each measurement. By  $K_{y_k}^{b_k}$  we denote the Kraus operators of Bob<sub>k</sub> associated to the  $y_k$ th measurement and  $b_k$ th outcome. The state received by Bob<sub>k</sub> is specified by Alice's input  $x$ , and the strings of inputs  $(y_1, \dots, y_{k-1})$  and outputs  $(b_1, \dots, b_{k-1})$  of all previous Bobs. However, we treat each Bob in the sequence as independent from the rest, meaning that they do not know the specific inputs or outputs of the other Bobs in each run of the experiment. Thus, in order to calculate the relevant marginal distributions  $p(b_k|x, y_k)$ , only the average state  $\tilde{\rho}_x^{(k)}$  received by Bob<sub>k</sub> is required, i.e., the state obtained from averaging a preparation  $\rho_x$  of Alice over all the inputs and outputs of all previous Bobs:

$$\tilde{\rho}_x^{(k)} = \frac{1}{n} \sum_{y_{k-1}, b_{k-1}} K_{y_{k-1}}^{b_{k-1}} \tilde{\rho}_x^{(k-1)} (K_{y_{k-1}}^{b_{k-1}})^\dagger, \quad (4)$$

with  $\tilde{\rho}_x^{(1)} = \rho_x$ . Consequently, the desired marginal statistics for Bob<sub>k</sub> are  $p(b_k|x, y_k) = \text{tr}(\tilde{\rho}_x^{(k)} (K_{y_k}^{b_k})^\dagger K_{y_k}^{b_k})$ . This constitutes a description of general quantum strategies in the sequential scenario.

### 3.1 Sharing preparation contextuality

We apply the above general description to construct a specific family of quantum strategies for sharing preparation contextuality, that is inspired by the previously described optimal quantum strategy for the maximal violation of the inequalities (1). Alice prepares the states (3) while each Bob performs an unsharp variant of the measurements optimal for violating (1). In that strategy the measurements of Bob are the dichotomic observables  $G_{n, y_k}^T$  defined in (2), corresponding to the projectors  $\Pi_{n, y}^b = (\mathbb{1} + (-1)^b G_{n, y}^T)/2$  that are both the Kraus operators and POVM elements. For a weaker measurement, one modifies the

POVM element to  $(\mathbb{1} + (-1)^b \eta_k G_{n, y}^T)/2$ , for some  $\eta_k \in [0, 1]$ . If  $\eta_k = 1$  ( $\eta_k = 0$ ), the measurement is sharp (non-interacting). Choosing  $0 < \eta_k < 1$  corresponds to an unsharp measurement. The corresponding Kraus operator is given by

$$K_{y_k}^{b_k} = \sqrt{\frac{1 + \eta_k}{2}} \Pi_{n, y_k}^{b_k} + \sqrt{\frac{1 - \eta_k}{2}} \Pi_{n, y_k}^{\bar{b}_k}, \quad (5)$$

where the bar-sign denotes a bit-flip. This class of strategies has the following convenient property.

**Lemma 1.** *If Alice prepares the states in Eq. (3) and the Bobs each measure  $G_{n, y_k}^T$  with sharpness  $\eta_k$ , the average state received by Bob<sub>k</sub> is*

$$\tilde{\rho}_x^{(k)} = v_k \rho_x + (1 - v_k) \rho_{\text{mix}}, \quad (6)$$

where  $\rho_{\text{mix}}$  is the maximally mixed state and the visibility  $v_k \in [0, 1]$  is given recursively by

$$v_k = v_{k-1} f_{k-1} = \prod_{j=1}^{k-1} f_j, \quad (7)$$

where  $v_1 = 1$  by definition, and the "quality factor"  $f_k$  of the measurement of Bob<sub>k</sub> is defined from the sharpness  $\eta_k$  as  $f_k = (1 + (n-1)\sqrt{1 - \eta_k^2})/n$ .

*Proof.* The proof is technical in character and is given in Appendix A. ■

Using Eq. (6), the figure of merit (1) for the pair Alice and Bob<sub>k</sub> reads

$$\mathcal{A}_k^{(n)} = \frac{1}{2} \left( 1 + \frac{v_k \eta_k}{\sqrt{n}} \right). \quad (8)$$

This leads to preparation contextuality whenever  $\eta_k > 1/(v_k \sqrt{n})$ . This can be used to recursively calculate the critical pairs  $(\eta_k, v_k)$ . Thusly, we arrive at the following result.

**Result 1.** *The number of observers who can independently share the preparation contextuality enabled by Alice's ensemble is at least  $n$ .*

*Proof.* Consider that each Bob tunes the sharpness of his measurement so as to just violate the inequality (1), but not more. Expressing the measurement sharpness  $\eta_k = \sin \theta_k$ , where  $\theta_k \in [0, \pi/2]$ , we thus require  $\sin \theta_k = 1/(v_k \sqrt{n})$ . On the other hand, a trivial lower bound on the quality factor of Bob<sub>k</sub>'s measurement is  $f_k = (1 + (n-1) \cos \theta_k)/n \geq \cos \theta_k$ . Squaring, and using the expression for the critical value of  $\sin \theta_k$  above, we find that  $f_k^2 \geq 1 - 1/(v_k^2 n)$ . Since the visibility of the next Bob is  $v_{k+1} = v_k f_k$ , we have  $v_{k+1}^2 = v_k^2 f_k^2 \geq v_k^2 (1 - 1/(v_k^2 n))$ . Hence, the decrease in visibility from each Bob to the next is bounded by  $v_k^2 - v_{k+1}^2 \leq 1/n$  which together with  $v_1 = 1$  gives  $v_{k+1}^2 \geq 1 - k/n$ . This implies that the visibility of the  $n$ th Bob is at least  $v_n \geq 1/\sqrt{n}$ , which is precisely the condition for violating the preparation noncontextuality inequality. ■

Thus by suitably choosing  $n$ , an arbitrary long sequence of observers can share the preparation contextual correlations enabled by Alice's ensemble. Moreover, we now show that for the considered class of quantum strategies, the number of observers who share preparation contextuality can be no more than  $n$ . Consider the quality factor  $f_k$  of the measurement of Bob $_k$ . We can find upper and lower bounds on  $f_k^2$  in the following manner. First, for a lower bound,

$$f_k = \frac{1 + (n-1) \cos \theta_k}{n} > \cos \theta_k, \quad (9)$$

$$\therefore f_k^2 > \cos^2 \theta_k = 1 - \sin^2 \theta_k = 1 - \frac{1}{nv_k^2}. \quad (10)$$

For the upper bound,

$$\begin{aligned} f_k^2 &< f_k^2 + 4 \frac{(n-1)}{n^2} \sin^4 \frac{\theta_k}{2} \\ &= 1 - \frac{(n-1)}{n} \sin^2 \theta_k = 1 - \frac{(n-1)}{n^2 v_k^2}. \end{aligned} \quad (11)$$

But since the visibility  $v_{k+1}$  of the next Bob is given by  $v_{k+1} = v_k f_k$ , we can bound the next visibility as

$$v_k^2 \left(1 - \frac{1}{nv_k^2}\right) < v_{k+1}^2 < v_k^2 \left(1 - \frac{(n-1)}{n^2 v_k^2}\right), \quad (12)$$

from which the decrease in the visibility squared is both bounded on both sides, by

$$\frac{1}{n} < v_k^2 - v_{k+1}^2 < \frac{n-1}{n^2}. \quad (13)$$

Proceeding from the first Bob, who has visibility  $v_1 = 1$ , we can use the lower bound to find that  $v_n^2 > 1/n$ , and the upper bound to find that  $v_{n+1}^2 < 1/n$ . Since  $1/\sqrt{n}$  is the critical visibility to violate the preparation noncontextuality inequality, it follows that Bob $_n$  can violate the inequality (as all of the Bobs before him), but that Bob $_{n+1}$  and later Bobs cannot.

Another noteworthy feature is that one can share preparation contextuality between any number of observers also in a scenario in which none of the Bob's knows his position in the sequence. To this end, consider a quantum strategy in which the set of possible measurements performed by each Bob is the same, i.e., they all perform equally unsharp measurements. If it is the case that the first  $k$  Bobs in the sequence violate the preparation noncontextuality inequality, then the weakest violation will be by the last Bob. The condition for the  $k$ 'th Bob to just saturate the preparation noncontextuality inequality reads

$$\sin \theta = \frac{1}{v_{k-1} \sqrt{n}}, \quad (14)$$

where  $\eta = \sin \theta$  is the strength of all of the Bobs' measurements, and the visibility is given by,

$$v_{k-1} = \left( \frac{1 + (n-1) \cos \theta}{n} \right)^{k-1}. \quad (15)$$

Solving the equation for the value of  $k$  returns

$$k = 1 - \frac{\log \sin \theta + \frac{1}{2} \log n}{\log(1 + (n-1) \cos \theta) - \log n}. \quad (16)$$

Consider that the strength of the measurement is chosen to be  $\sin \theta = \sqrt{e/n}$  (where  $n \geq 3$ ). In any case, we are interested in the scaling for large  $n$ , for which we may approximate  $\cos \theta = \sqrt{1 - e/n} \approx 1 - e/2n$ . Substituting this in the above, and further approximating  $\log(1-x) \approx -x$  for  $|x| \ll 1$ , one gets

$$k \approx 1 + \frac{n^2}{(n-1)e} = O\left(\frac{n}{e}\right). \quad (17)$$

One can show that this is the optimal scaling by the Maclaurin expansion of (16) for small  $\theta$ , and differentiating to find the optimal value of  $\theta$ .

Thus we find that even in the anonymous setting where each Bob is unaware of their position in the sequence, the maximum number of observers able to share the contextuality enabled by Alice's ensemble by all performing equally unsharp measurements scales as

$$k_{\max} \approx \frac{n}{e}. \quad (18)$$

Note that this scaling is the same as obtained in the non-anonymous setting, up to a pre-factor of  $1/e$ .

### 3.2 Noise-robustness

The scenario we have considered so far is an idealisation in which no noise appears. In addition to this not being realistic in any experiment, it is interesting to consider whether the noiseless scenario is distinctive, or also significantly noisy ensembles [36] enable shared preparation contextuality. To address this matter, we let Alice's preparations be mixtures of the intended state  $\rho_x$  with the maximally mixed state:  $\rho_x(q) = q\rho_x + (1-q)\rho_{\text{mix}}$  for some visibility  $q \in [0, 1]$ . For a given number of observers, what is the smallest  $q$  such that preparation contextuality can be shared between all observers?

**Result 2.** *For any given number of independent observers  $m$ , there exists an ensemble whose contextuality can be shared between all observers for any  $q > 0$ .*

*Proof.* The proof follows the same steps as that of Result 1. We substitute  $\rho_x$  for  $\rho_x(q)$  in the proof of Result 1, which means that  $v_1 = q$ . Following the same procedure, one obtains the following condition for the visibility in the sequence:  $v_{k+1}^2 \geq q - k/n$ . This implies that to find a violation for the  $m$ 'th Bob, one must choose a sufficiently large  $n$ , namely  $n \geq \lceil \frac{m}{q} \rceil$ . ■

Hence, preparation contextuality can be shared between any number of observers using ensembles with an arbitrarily large noise-component by choosing a sufficiently large  $n$ . The price to pay for this property is that when  $q \rightarrow 0$ , both the Hilbert space dimension



of Alice’s ensemble and the number of preparations and measurements diverge.

## 4 Experiment

We demonstrate the theoretical findings in an experiment with three ( $n = 3$ ) sequential tests of preparation contextuality. Alice prepares the eight qubit states (3) with Bloch vectors  $\vec{a}_x = [(-1)^{x_1}, (-1)^{x_2}, (-1)^{x_3}] / \sqrt{3}$ . Bob<sub>1</sub> and Bob<sub>2</sub> perform unsharp measurements (5) of  $\sigma_x$ ,  $\sigma_y$  and  $\sigma_z$  whereas Bob<sub>3</sub> performs projective (sharp) measurements of the same observables.

In the experiment we perform unsharp measurements on the polarisation state of a single photon using shifted Sagnac interferometers, as shown in Bob<sub>1</sub> and Bob<sub>2</sub> in Fig. (2). A HWP is placed in each path of the interferometer, rotated to  $\theta_i/2$  in the horizontal path and  $\pi/4 - \theta_i/2$  in the vertical path to control the sharpness of the measurement. A HWP and QWP before and after the interferometer are used to select the basis of the measurement. The measurement outcome is encoded in the output path, i.e. outcome  $b_i = 0$  ( $b_i = 1$ ) corresponds to the detection of the photon in output path 1 (2, beam blocked in figure). In the sequential scenario we choose to consider only one path at a time for Bob<sub>1</sub> and Bob<sub>2</sub> to simplify the set-up. By adding an additional rotation to the HWPs or QWPs before and after Bob, we can select the output we want to analyse [10, 11]. The results of Bob<sub>1</sub> and Bob<sub>2</sub>’s unsharp measurements are therefore obtained at Bob<sub>3</sub>, comprised of a PBS and single photon detectors D<sub>1</sub> and D<sub>2</sub>. For example, if we consider output 1 at Bob<sub>1</sub> and Bob<sub>2</sub>, a click in either detector at Bob<sub>3</sub> tells us that Bob<sub>1</sub> and Bob<sub>2</sub> had the outcome  $b_1 = 0$  and  $b_2 = 0$ . This is, however, not a limitation since we can equally well have Bob<sub>1</sub> and Bob<sub>2</sub> read out their respective outputs by placing an additional single photon detector in their respective labs to detect the photon path. We analyse the counts in Bob<sub>3</sub> corresponding to all possible combinations of output ports to realise a full measurement. This protocol relies on a stable photon generation rate. Details of measurement angles are given in Appendix B. This set-up can be used to perform projective measurements ( $\eta = 1$ ,  $\theta_i = 0$ ), no measurement ( $\eta = 0$ ,  $\theta_i = \pi/4$ ), or an intermediate-strength measurement, where the sharpness (strength) of the measurement is tuned by varying  $\theta_i$ .

The full set-up is shown in Fig. 2. We generate heralded single photons at 780 nm via spontaneous parametric down-conversion (SPDC) using a single type-I beta barium borate (BBO) crystal of thickness 2 mm pumped by 390 nm femto-second laser pulses. The idler photon is detected by an APD single-photon detector, D<sub>trigger</sub>, and is used as a trigger. The single photons are coupled into single-mode fibres (SMF) after passing through a narrowband 3 nm interference

filter (F) to define the spatial and spectral properties of the photons. After filtering, the signal photon is prepared into one of Alice’s eight states, using a polariser, two QWPs and a HWP (angles given in Appendix B). The unsharp measurements of Bob<sub>1</sub> and Bob<sub>2</sub> correspond to  $\theta_1 = 24.95^\circ$  ( $\eta_1 = 0.6441$ ) and  $\theta_2 = 20.10^\circ$  ( $\eta_2 = 0.7637$ ) respectively, which ideally produce  $\mathcal{A}_1 = \mathcal{A}_2 = \mathcal{A}_3 = 0.6859 > 2/3$  with  $\mathcal{A}_k = \mathcal{A}_k^{(3)}$ .

### 4.1 Results

In order to test each of the three preparation non-contextuality inequalities (between Alice and each of the three Bobs), we require 24 marginal probabilities (the ‘winning’ answers  $b_k = x_{y_k}$ ) corresponding to the three measurement bases and Alice’s eight preparations. To reduce the Poissonian error, each Bob collects approximately 34 million counts for each of these 24 settings. Our experimental values can be found in Appendix C. These lead to three preliminary values of  $\mathcal{A}_1^{\text{pre}} = 0.687 \pm 0.001$ ,  $\mathcal{A}_2^{\text{pre}} = 0.675 \pm 0.001$ , and  $\mathcal{A}_3^{\text{pre}} = 0.681 \pm 0.001$ .

### 4.2 Data analysis

Due to small yet unavoidable experimental imperfections, e.g. waveplate imperfections and offsets in the rotation of the waveplates, it is impossible to perfectly satisfy the operational indistinguishability relations required to test preparation contextuality. This problem can be overcome by suitable post-processing methods [30]. As described in Appendix D, we have used a relaxed variant of these methods to enforce the indistinguishability relations relevant to a test of inequality (1) on our experimental data. This comes at the cost of the observed values ( $\mathcal{A}_1^{\text{pre}}$ ,  $\mathcal{A}_2^{\text{pre}}$ ,  $\mathcal{A}_3^{\text{pre}}$ ) decreasing in a manner corresponding to how well the statistics approximates said relations. Due to the high visibility and precision of the experimental set-up, we find only a small decrease in the three correlation witnesses:

$$\mathcal{A}_1^{\text{post}} = 0.683 \pm 0.001$$

$$\mathcal{A}_2^{\text{post}} = 0.670 \pm 0.001$$

$$\mathcal{A}_3^{\text{post}} = 0.677 \pm 0.001$$

all of which violate inequality (1).

## 5 Conclusions

We have theoretically developed and experimentally demonstrated the sharing of preparation contextual correlations in scenarios that require no entanglement. In addition to such correlations being possible to share between any number of observers, we found

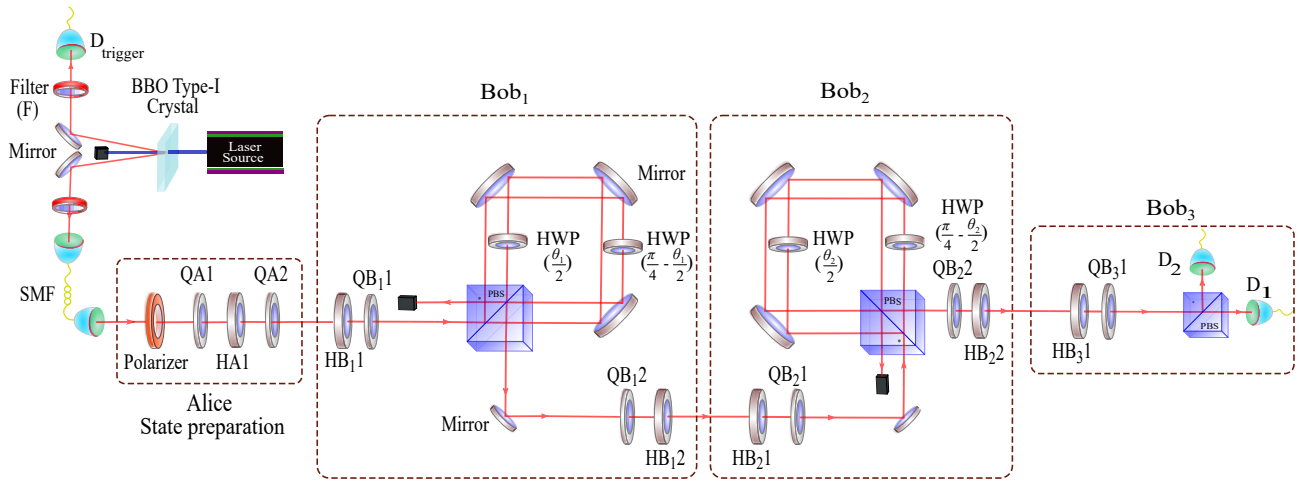


Figure 2: Optical set-up used to reveal contextuality sharing. See text for details. Q and H represent quarter-wave plates (QWPs) and half-wave plates (HWP).

that this can be done in a strongly noise-robust manner. This distinguishes shared preparation contextuality from known results in e.g. shared Bell nonlocality in which the fragility to noise of sequential demonstrations scales super-exponentially [4]. This fragility poses a significant experimental hurdle and has hitherto limited demonstrations to two sequential violations of Bell inequalities [10, 11]. We experimentally observed three sequential demonstrations of preparation contextuality. Optical set-ups of this spirit (see also Refs [10, 11]) are promising candidates for a variety of sequential correlation tests. Finally, an interesting question is to understand which forms of quantum correlations can be shared between indefinitely many observers in a noise-robust manner.

*Note added.*— In the substantial time between the arXiv submission of this work and its publication, also other works concerning the sequential sharing of contextuality have appeared [28, 37].

## 6 Acknowledgements

H. A. and N. W. contributed equally to this work. This work was supported by the project “Photonic Quantum Information” (Knut and Alice Wallenberg Foundation, Sweden), the Swedish Research Council, Stiftelsen Olle Engkvist Byggmästare and the Swiss National Science Foundation (Starting grant DIAQ, Early PostDoc Mobility fellowship P2GEP2 194800, and the National Centre of Competence in Research “Quantum Systems and Technology”).

## References

- [1] C. A. Fuchs, and A. Peres, Quantum-state disturbance versus information gain: Uncertainty relations for quantum information *Phys. Rev. A* **53**, 2038 (1996).
- [2] R. Gallego, L. E. Würflinger, R. Chaves, A. Acín, M. Navascués, Nonlocality in sequential correlation scenarios, *New J. Phys.* **16**, 033037 (2014).
- [3] C. Budroni, T. Moroder, M. Kleinmann, and O. Gühne, Bounding Temporal Quantum Correlations, *Phys. Rev. Lett.* **111**, 020403 (2013).
- [4] R. Silva, N. Gisin, Y. Guryanova, and S. Popescu, Multiple Observers Can Share the Nonlocality of Half of an Entangled Pair by Using Optimal Weak Measurements, *Phys. Rev. Lett.* **114**, 250401 (2015).
- [5] A. Tavakoli, A. Cabello, Quantum predictions for an unmeasured system cannot be simulated with a finite-memory classical system, *Phys. Rev. A* **97**, 032131 (2018).
- [6] P. J. Brown and R. Colbeck Arbitrarily Many Independent Observers Can Share the Nonlocality of a Single Maximally Entangled Qubit Pair, *Phys. Rev. Lett.* **125**, 090401 (2020).
- [7] F. J. Curchod, M. Johansson, R. Augusiak, M. J. Hoban, P. Wittek, and A. Acín, Unbounded randomness certification using sequences of measurements, *Phys. Rev. A* **95**, 020102(R) (2017).
- [8] B. Coyle, M. J. Hoban, and E. Kashefi, One-Sided Device-Independent Certification of Unbounded Random Numbers, *EPTCS* **273**, 14-26 (2018).
- [9] G. Foletto, L. Calderaro, A. Tavakoli, M. Schiavon, F. Picciariello, A. Cabello, P. Villoresi, and G. Vallone, Experimental Certification of Sustained Entanglement and Nonlocality after Sequential Measurements, *Phys. Rev. Applied* **13**, 044008 (2020).
- [10] M. Schiavon, L. Calderaro, M. Pittaluga, G. Vallone, and P. Villoresi, Three-observer Bell inequality violation on a two-qubit entangled state, *Quantum Sci. Technol.* **2** 015010 (2017).
- [11] M-J. Hu, Z-Y. Zhou, X-M. Hu, C-F. Li, G-C. Guo, and Y-S. Zhang, Observation of non-

- locality sharing among three observers with one entangled pair via optimal weak measurement, [npj Quantum Information](#) **4**, 63 (2018).
- [12] A. Bera, S. Mal, A. Sen De, and U. Sen, Witnessing bipartite entanglement sequentially by multiple observers, [Phys. Rev. A](#) **98**, 062304 (2018).
- [13] S. Sasmal, D. Das, S. Mal, and A.S. Majumdar, Steering a single system sequentially by multiple observers, [Phys. Rev. A](#) **98**, 012305 (2018).
- [14] A. Shenoy H, S. Designolle, F. Hirsch, R. Silva, N. Gisin, and N. Brunner, Unbounded sequence of observers exhibiting Einstein-Podolsky-Rosen steering, [Phys. Rev. A](#) **99**, 022317 (2019).
- [15] K. Mohan. A. Tavakoli, and N. Brunner, Sequential random access codes and self-testing of quantum measurement instruments, [New J. Phys.](#) **21** 083034 (2019).
- [16] N. Miklin, J. Borkala, and M. Pawłowski, Semi-device-independent self-testing of unsharp measurements, [Phys. Rev. Research](#) **2**, 033014 (2020)
- [17] H. Anwer, S. Muhammad, W. Cherifi, N. Miklin, A. Tavakoli, and M. Bourennane, Experimental Characterization of Unsharp Qubit Observables and Sequential Measurement Incompatibility via Quantum Random Access Codes, [Phys. Rev. Lett.](#) **125**, 080403 (2020).
- [18] G. Foletto, L. Calderaro, G. Vallone, and P. Villoresi, Experimental demonstration of sequential quantum random access codes, [Phys. Rev. Research](#) **2**, 033205 (2020).
- [19] R. W. Spekkens, Contextuality for preparations, transformations, and unsharp measurements [Phys. Rev. A](#) **71**, 052108 (2005).
- [20] R. W. Spekkens, Negativity and Contextuality are Equivalent Notions of Nonclassicality, [Phys. Rev. Lett.](#) **101**, 020401 (2008).
- [21] R. W. Spekkens, D. H. Buzacott, A. J. Keehn, B. Toner, and G. J. Pryde, Preparation Contextuality Powers Parity-Oblivious Multiplexing [Phys. Rev. Lett.](#) **102**, 010401 (2009).
- [22] M. S. Leifer, and O. J. E. Maroney, Maximally Epistemic Interpretations of the Quantum State and Contextuality, [Phys. Rev. Lett.](#) **110**, 120401 (2013).
- [23] M. Banik, S. S. Bhattacharya, A. Mukherjee, A. Roy, A. Ambainis, and A. Rai, Limited preparation contextuality in quantum theory and its relation to the Cirel'son bound, [Phys. Rev. A](#) **92**, 030103(R) (2015).
- [24] S. Ghorai, A. K. Pan, Optimal quantum preparation contextuality in an n-bit parity-oblivious multiplexing task [Phys. Rev. A](#) **98**, 032110 (2018).
- [25] A. Tavakoli and R. Uola, Measurement incompatibility and steering are necessary and sufficient for operational contextuality, [Phys. Rev. Research](#) **2**, 013011 (2020).
- [26] D. Saha, and A. Chaturvedi, Preparation contextuality: the ground of quantum communication advantage, [Phys. Rev. A](#) **100**, 022108 (2019).
- [27] A. Tavakoli, E. Zambrini Cruzeiro, R. Uola, and A. A. Abbott, Bounding and Simulating Contextual Correlations in Quantum Theory, [PRX Quantum](#) **2**, 020334 (2021).
- [28] A. Chaturvedi, M. Farkas, and V. Wright, Characterising and bounding the set of quantum behaviours in contextuality scenarios, [Quantum](#) **5**, 484 (2021).
- [29] A. Hameedi, A. Tavakoli, B. Marques, and M. Bourennane, Communication games reveal preparation contextuality, [Phys. Rev. Lett.](#) **119**, 220402 (2017)
- [30] M. D. Mazurek, M. F. Pusey, R. Kunjwal, K. J. Resch, and R. W. Spekkens, An experimental test of noncontextuality without unphysical idealizations, [Nature Communications](#) **7**, 11780 (2016).
- [31] S. Kochen, and E. P. Specker, The Problem of Hidden Variables in Quantum Mechanics, [Indiana University Mathematics Journal](#), **17**, 59 (1967).
- [32] N. Harrigan, and R. W. Spekkens, Einstein, Incompleteness, and the Epistemic View of Quantum States, [Found Phys](#) (2010) **40**, 125 (2010).
- [33] A. Ambainis, A. Nayak, A. Ta-Shma, U. Vazirani, Dense quantum coding and a lower bound for 1-way quantum automata, [Proceedings of the 31st Annual ACM Symposium on Theory of Computing \(STOC'99\)](#), 376-383 (1999).
- [34] A. Tavakoli, A. Hameedi, B. Marques, and M. Bourennane, Quantum random access codes using single d-Level systems, [Phys. Rev. Lett.](#) **114**, 170502 (2015).
- [35] A. Chailloux, I. Kerenidis, S. Kundu, and J. Sikora, Optimal bounds for parity-oblivious random access codes, [New J. Phys.](#) **18**, 045003 (2016).
- [36] One could alternatively consider the Bobs' measurement devices inducing the noise. However, this is less detrimental than noisy preparations. The reason is that if Alice's preparations are noisy the correlations due to all Bobs' measurements are weaker, whereas if instead one (or many) of the Bobs sometimes fail to perform the intended measurement, the state relayed to the next Bob retains a higher degree of coherence and leads to him observing stronger correlations.
- [37] A. Kumari and A. K. Pan, Sharing nonlocality and nontrivial preparation contextuality using the same family of Bell expressions, [Phys. Rev. A](#) **100**, 062130 (2019).

## A Proof of Lemma

In this section, we prove the lemma of the main text. In the considered scenario, Alice receives a random input  $x \in \{0, 1\}^n$  and prepares the associated state

$$\rho_x = \text{tr}_A \left[ (\mathbb{1} + A_x) \otimes \mathbb{1} (|\phi_{\max}\rangle\langle\phi_{\max}|)^{\otimes \lfloor n/2 \rfloor} \right], \quad (19)$$

where  $(|\phi_{\max}\rangle\langle\phi_{\max}|)^{\otimes \lfloor n/2 \rfloor}$  is  $\lfloor n/2 \rfloor$  copies of the two-qubit maximally entangled state, and the partial trace is taken over all the first qubits in each pair. Consider that the sequence of Bobs, labelled by  $\{1, 2, \dots, m-1\}$ , apply measurements of intermediate sharpness to the state above, each denoted by  $\eta_k = \sin \theta_k$ . We proceed to prove that the average state  $\tilde{\rho}_x^{(m)}$  received by Bob $_m$  will be of the form

$$\tilde{\rho}_x^{(m)} = \text{tr}_A \left[ (\mathbb{1} + v_m A_x) \otimes \mathbb{1} (|\phi_{\max}\rangle\langle\phi_{\max}|)^{\otimes \lfloor n/2 \rfloor} \right], \quad (20)$$

where  $v_m$  (the ‘‘visibility’’ of the state) is given by

$$v_m = v_{m-1} f_{m-1} = \prod_{j=1}^{m-1} f_j, \quad (21)$$

$$\text{where } f_j = \frac{1 + (n-1) \cos \theta_j}{n}. \quad (22)$$

We call  $f_j$  the ‘‘quality factor’’ of the measurement of the  $j^{\text{th}}$  Bob. The visibility of the first Bob is  $v_1 = 1$ , since he possesses the undisturbed state received directly from Alice.

The proof is inductive. For the first Bob, the statement holds trivially. Consider that it holds true for  $m-1$  Bobs, so that the average state  $\tilde{\rho}_x^{(m)}$  received by Bob $_m$  is given by (20). Then using the Kraus operators stated in the main text, the average state  $\tilde{\rho}_x^{(m+1)}$  (averaging over all Bob $_m$ 's possible and equiprobable inputs, and with no knowledge of his outcome), is given by

$$\tilde{\rho}_x^{(m+1)} = \frac{1}{n} \sum_{y,b} K_y^b \tilde{\rho}_x^{(m)} (K_y^b)^\dagger = \frac{1}{n} \sum_{y,b} \text{tr}_A \left[ (\mathbb{1} + v_m A_x) \otimes K_y^b (|\phi_{\max}\rangle\langle\phi_{\max}|)^{\otimes \lfloor n/2 \rfloor} \mathbb{1} \otimes (K_y^b)^\dagger \right], \quad (23)$$

where the Kraus operators are acting on the part of the Hilbert space complementary to that being traced out. First, using the property of the maximally entangled state that  $(\mathbb{1} \otimes O) |\phi_{\max}\rangle\langle\phi_{\max}| (\mathbb{1} \otimes O^\dagger) = (O^T \otimes \mathbb{1}) |\phi_{\max}\rangle\langle\phi_{\max}| (O^* \otimes \mathbb{1})$ , and then using the cyclicity of the trace, we obtain

$$\tilde{\rho}_x^{(m+1)} = \frac{1}{n} \sum_{y,b} \text{tr}_A \left[ (\mathbb{1} + v_m A_x) (K_y^b)^T \otimes \mathbb{1} (|\phi_{\max}\rangle\langle\phi_{\max}|)^{\otimes \lfloor n/2 \rfloor} (K_y^b)^\dagger \otimes \mathbb{1} \right] \quad (24)$$

$$= \frac{1}{n} \sum_{y,b} \text{tr}_A \left[ (K_y^b)^\dagger \otimes (\mathbb{1} + v_m A_x) (K_y^b)^T \otimes \mathbb{1} (|\phi_{\max}\rangle\langle\phi_{\max}|)^{\otimes \lfloor n/2 \rfloor} \right]. \quad (25)$$

Splitting the above into the sum of the two terms from the  $(\mathbb{1} + v_m A_x)$ , the contribution of the  $\mathbb{1}$  part is

$$\begin{aligned} \frac{1}{n} \sum_{y,b} \text{tr}_A \left[ (K_y^b)^\dagger \otimes (K_y^b)^T \otimes \mathbb{1} (|\phi_{\max}\rangle\langle\phi_{\max}|)^{\otimes \lfloor n/2 \rfloor} \right] &= \frac{1}{n} \sum_y \text{tr}_A \left[ \mathbb{1} \otimes \mathbb{1} (|\phi_{\max}\rangle\langle\phi_{\max}|)^{\otimes \lfloor n/2 \rfloor} \right] \\ &= \text{tr}_A \left[ (|\phi_{\max}\rangle\langle\phi_{\max}|)^{\otimes \lfloor n/2 \rfloor} \right], \end{aligned} \quad (26)$$

where we have used that the  $K_y^b$  are Hermitian and that measurements are complete i.e.,  $\sum_b (K_y^b)^\dagger (K_y^b)^T = \mathbb{1}$ .

For the term involving  $A_x$ , we calculate the sum using the Kraus operators from the main text, denoting by  $\eta_m = \sin \theta_m$  the strength of the measurement of Bob  $m$ ,

$$K_y^b = \sqrt{\frac{1 + \eta_m}{2}} \Pi_{n,y}^b + \sqrt{\frac{1 - \eta_m}{2}} \Pi_{n,y}^{\bar{b}} = \left( \frac{\cos \frac{\theta_m}{2} \mathbb{1} + (-1)^b \sin \frac{\theta_m}{2} G_{n,y}^T}{\sqrt{2}} \right), \quad (27)$$



which results in

$$\begin{aligned}
\frac{1}{n} \sum_{y,b} (K_y^b)^{\dagger T} A_x (K_y^b)^T &= \frac{1}{n} \sum_{y,b} \left( \frac{\cos \frac{\theta_m}{2} \mathbb{1} + (-1)^b \sin \frac{\theta_m}{2} G_{n,y}}{\sqrt{2}} \right) A_x \left( \frac{\cos \frac{\theta_m}{2} \mathbb{1} + (-1)^b \sin \frac{\theta_m}{2} G_{n,y}}{\sqrt{2}} \right) \\
&= \frac{1}{2n} \sum_{y,b} \cos^2 \left( \frac{\theta_m}{2} \right) A_x + (-1)^b \cos \left( \frac{\theta_m}{2} \right) \sin \left( \frac{\theta_m}{2} \right) \{G_{n,y}, A_x\} + \sin^2 \left( \frac{\theta_m}{2} \right) G_{n,y} A_x G_{n,y}. \\
&= \frac{1}{n} \sum_y \left( \frac{1 + \cos \theta_m}{2} \right) A_x + \left( \frac{1 - \cos \theta_m}{2} \right) G_{n,y} A_x G_{n,y} \\
&= \left( \frac{1 + \cos \theta_m}{2} \right) A_x + \left( \frac{1 - \cos \theta_m}{2} \right) \frac{1}{n} \sum_y G_{n,y} A_x G_{n,y}. \tag{28}
\end{aligned}$$

We may now use the expansion  $A_x = \frac{1}{\sqrt{n}} \sum_i (-1)^{x_i} G_{n,i}$ , and the anti-commutation relation  $\{G_{n,j}, G_{n,k}\} = 2\delta_{j,k} \mathbb{1}$  from [35] to simplify

$$\begin{aligned}
\frac{1}{n} \sum_y G_{n,y} A_x G_{n,y} &= \frac{1}{\sqrt{n}} \sum_i (-1)^{x_i} \frac{1}{n} \sum_y G_{n,y} G_{n,i} G_{n,y} \\
&= \frac{1}{\sqrt{n}} \sum_i (-1)^{x_i} \frac{1}{n} \sum_y (2\delta_{i,y} G_{n,y} - G_{n,i}) \\
&= \frac{1}{\sqrt{n}} \sum_i (-1)^{x_i} \frac{1}{n} (2 - n) G_{n,i} \\
&= \frac{2 - n}{n} A_x. \tag{29}
\end{aligned}$$

Inserting this into Eq. (28), we obtain

$$\frac{1}{n} \sum_{y,b} (K_y^b)^{\dagger T} A_x (K_y^b)^T = f_m A_x, \tag{30}$$

$$\text{where } f_m = \left( \frac{1 + (n-1) \cos \theta_m}{n} \right) = \left( \frac{1 + (n-1) \sqrt{1 - \eta^2}}{n} \right), \tag{31}$$

is the quality factor of the measurement of Bob<sub>m</sub>. Combining this with Eq. (26) to find the final expression for the average state after Bob<sub>m</sub>'s measurement Eq. (25), we find

$$\tilde{\rho}_x^{(m+1)} = \text{tr}_A \left[ (|\phi_{\max}\rangle\langle\phi_{\max}|)^{\otimes [n/2]} \right] + \text{tr}_A \left[ v_m f_m A_x \otimes \mathbb{1} (|\phi_{\max}\rangle\langle\phi_{\max}|)^{\otimes [n/2]} \right] \tag{32}$$

$$= \text{tr}_A \left[ (\mathbb{1} + v_m f_m A_x) \otimes \mathbb{1} (|\phi_{\max}\rangle\langle\phi_{\max}|)^{\otimes [n/2]} \right], \tag{33}$$

which proves the desired relation ((20) - (22)).

## B Experimental settings

The angles used for Alice's state preparation are given below:

The settings of the HWPs and QWPs used for the unsharp measurements in Bob<sub>1</sub> and Bob<sub>2</sub> are as follows. Note these settings are independent of the sharpness of the measurement, which is determined by the angle of the HWPs inside the interferometer.

State	Pol. (°)	QA1 (°)	HA1 (°)	QA2 (°)
000	27.37	45	-33.75	45
001	27.37	45	-11.25	45
010	27.37	45	-56.25	45
011	27.37	45	-78.75	45
100	62.63	45	-33.75	45
101	62.63	45	-11.25	45
110	62.63	45	-56.25	45
111	62.63	45	-78.75	45

Table 1: Angles for the polarizer, QWPs and HWP for the preparation of Alice's states.

Measurement	Output Port	HB <sub>i</sub> 1 (°)	QB <sub>i</sub> 1 (°)	HB <sub>i</sub> 2(°)	QB <sub>i</sub> 2 (°)
$\sigma_x$	1	22.5	0	90	22.5
$\sigma_x$	2	67.5	0	90	67.5
$\sigma_y$	1	0	-45	45	0
$\sigma_y$	2	0	45	135	0
$\sigma_z$	1	0	0	90	0
$\sigma_z$	2	45	0	90	45

Table 2: Waveplate settings for measurement and output selection of Bob<sub>1</sub> and Bob<sub>2</sub>.

## C Experimental results

The experimental marginal probabilities corresponding to the outcomes that satisfy  $b_i = x_{y_i}$  (the ‘winning’ answer in the communication game) for Bob<sub>1</sub> and Bob<sub>2</sub>’s unsharp measurements and Bob<sub>3</sub>’s projective measurements of  $\sigma_x$ ,  $\sigma_y$  and  $\sigma_z$  on each of Alice’s preparations are shown in the following three tables:

State	Bob <sub>1</sub>		
	$\sigma_x$	$\sigma_y$	$\sigma_z$
000	0.7369 ± 0.0003	0.7044 ± 0.0003	0.6593 ± 0.0002
001	0.6473 ± 0.0002	0.7257 ± 0.0003	0.7079 ± 0.0003
010	0.6900 ± 0.0003	0.6727 ± 0.0002	0.6571 ± 0.0002
011	0.6879 ± 0.0003	0.6501 ± 0.0002	0.7005 ± 0.0003
100	0.6911 ± 0.0003	0.6195 ± 0.0002	0.7180 ± 0.0003
101	0.6813 ± 0.0003	0.6464 ± 0.0002	0.6779 ± 0.0003
110	0.6400 ± 0.0002	0.7471 ± 0.0003	0.7125 ± 0.0003
111	0.7242 ± 0.0003	0.7132 ± 0.0003	0.6755 ± 0.0002

Table 3: Experimental marginal probabilities for Bob<sub>1</sub>.

State	Bob <sub>2</sub>		
	$\sigma_x$	$\sigma_y$	$\sigma_z$
000	$0.6997 \pm 0.0003$	$0.6422 \pm 0.0002$	$0.6851 \pm 0.0003$
001	$0.6586 \pm 0.0002$	$0.6785 \pm 0.0002$	$0.6746 \pm 0.0002$
010	$0.6537 \pm 0.0002$	$0.7088 \pm 0.0003$	$0.6715 \pm 0.0002$
011	$0.6896 \pm 0.0003$	$0.6824 \pm 0.0003$	$0.6572 \pm 0.0002$
100	$0.7106 \pm 0.0003$	$0.6370 \pm 0.0002$	$0.6775 \pm 0.0002$
101	$0.6446 \pm 0.0002$	$0.6792 \pm 0.0003$	$0.6868 \pm 0.0003$
110	$0.6553 \pm 0.0002$	$0.7000 \pm 0.0003$	$0.6752 \pm 0.0002$
111	$0.6787 \pm 0.0002$	$0.6666 \pm 0.0002$	$0.6853 \pm 0.0002$

Table 4: Experimental marginal probabilities for Bob<sub>2</sub>.

State	Bob <sub>3</sub>		
	$\sigma_x$	$\sigma_y$	$\sigma_z$
000	$0.7044 \pm 0.0003$	$0.6470 \pm 0.0002$	$0.6786 \pm 0.0003$
001	$0.6661 \pm 0.0002$	$0.6886 \pm 0.0003$	$0.6854 \pm 0.0003$
010	$0.6582 \pm 0.0002$	$0.7113 \pm 0.0003$	$0.6702 \pm 0.0002$
011	$0.7011 \pm 0.0003$	$0.6783 \pm 0.0003$	$0.6746 \pm 0.0003$
100	$0.6975 \pm 0.0003$	$0.6558 \pm 0.0002$	$0.6915 \pm 0.0003$
101	$0.6655 \pm 0.0003$	$0.7049 \pm 0.0003$	$0.6891 \pm 0.0003$
110	$0.6469 \pm 0.0002$	$0.6942 \pm 0.0003$	$0.6881 \pm 0.0003$
111	$0.7027 \pm 0.0003$	$0.6512 \pm 0.0002$	$0.6853 \pm 0.0003$

Table 5: Experimental marginal probabilities for Bob<sub>3</sub>.

## D Enforcing strict operational equivalences on experimental data

Tests of preparation contextuality require that the observed probabilities satisfy an equivalence relation. In the specific preparation noncontextuality inequalities considered in the main text, that equivalence relation follows from the indistinguishability relation imposed on Alice’s quantum preparations, i.e. that she hides the value of the parity  $r \cdot x$  for every string  $r \in \{0, 1\}^n$  with  $|r| \geq 2$ . This is an operational equivalence relation that is expressed in terms of probabilities as follows,

$$\forall r, \forall M : \sum_{r \cdot x=0} p(P_x|b, M) = \sum_{r \cdot x=1} p(P_x|b, M). \quad (34)$$

Evidently, due to unavoidable experimental imperfections, such a constraint can never be exactly satisfied. This necessitates data processing methods to contend with the problem. Ref. [30] developed a method for post-processing outcome statistics that approximately satisfies an operational equivalence constraint into data that strictly satisfies said constraint. The price to pay for this mapping is that the value of the witness after post-processing is worse than what is originally measured. Roughly speaking, the closer the unprocessed outcome statistics is to satisfying the operational equivalence constraint, the smaller the decrease in the witness value due to the post-processing scheme.

We have applied a simplified variant (which assumes that the experiment is accurately described by quantum theory) of the method of [30] to enforce operational equivalence in each of the three sequential tests of preparation contextuality. We describe how it applies to the experimental results of the pair Alice-Bob<sub>1</sub>. Since the the outcomes are binary, the full distribution  $p(b_1|x, y_1)$  can be described by only considering  $p(b_1 = 0|x, y_1)$ . We can write this distribution as eight vectors  $\mathbf{P}_x = [p(0|x, 1), p(0|x, 2), p(0|x, 3)]$ . The vectors  $\mathbf{P}_x$  will not perfectly satisfy the operational equivalence constraint (34). Therefore, we aim to map them to other distributions  $\mathbf{P}'_x$  which perfectly satisfy (34). This can be done by noting that an experiment in which  $\{\mathbf{P}_x\}$  is realised, also constitutes an effective realisation of all distributions in the convex hull of  $\{\mathbf{P}_x\}$  (due to linearity). Hence, we set

$$\mathbf{P}'_x = \sum_{x'} \omega_x^{x'} \mathbf{P}_{x'}, \quad (35)$$

where for  $\forall x$   $\{\omega_x^{x'}\}_{x'}$  is a probability distribution. We search a set of distributions  $\{\omega_x\}$  that maximises the witness of preparation contextuality while also enforcing (34). This problem is solved with a linear program

$$\mathcal{A}_1^{\text{post}} = \max_{\{\omega\}} \mathcal{A}_1^{\text{pre}}[\{\mathbf{P}'_x\}] \quad \text{such that } \forall r \in \{011, 101, 110, 111\} \quad \sum_{r \cdot x=0} \mathbf{P}'_x = \sum_{r \cdot x=1} \mathbf{P}'_x. \quad (36)$$

In addition, we can employ the quantity  $F = \sum_x \omega_x^x$  as a measure of the closeness of the observed and post-processed data. Moreover, this procedure can be straightforwardly adapted to the experimental results obtained for Alice-Bob<sub>2</sub> and Alice-Bob<sub>3</sub>. The minor difference is that the preparation procedure for e.g. Alice-Bob<sub>2</sub> effectively becomes the average state relayed by Bob<sub>1</sub> to Bob<sub>2</sub>. Thus, change the definition of vectors  $\mathbf{P}_x$  to instead apply to the distributions  $p(b_2 = 0|x, y_2)$  and  $p(b_3 = 0|x, y_3)$  respectively and proceed in analogy with the above.

Solving the above linear program, we have obtained the following results for the three demonstrations of preparation contextuality.

$$\mathcal{A}_1^{\text{pre}} = 0.687 \quad \mathcal{A}_1^{\text{post}} = 0.683 \quad F = 0.9690 \quad (37)$$

$$\mathcal{A}_2^{\text{pre}} = 0.675 \quad \mathcal{A}_2^{\text{post}} = 0.670 \quad F = 0.9537 \quad (38)$$

$$\mathcal{A}_3^{\text{pre}} = 0.681 \quad \mathcal{A}_3^{\text{post}} = 0.677 \quad F = 0.9700 \quad (39)$$