Alternative robust ways of witnessing nonclassicality in the simplest scenario

Massy Khoshbin [1,*] Lorenzo Catani [2,†] and Matthew Leifer[3]

[1]*Department of Mathematics, University of California, Santa Barbara, California 93106, USA*

[2]*International Iberian Nanotechnology Laboratory, Av. Mestre José Veiga s/n, 4715-330 Braga, Portugal*

[3]*Institute for Quantum Studies and Schmid College of Science and Technology, Chapman University, One University Drive, Orange, California 92866, USA*

(Received 27 November 2023; accepted 20 February 2024; published 14 March 2024)

In this paper we relate notions of nonclassicality in what is known as the simplest nontrivial scenario (a prepare and measure scenario composed of four preparations and two binary-outcome tomographically complete measurements). Specifically, we relate the established method developed by Pusey [M. F. Pusey, Phys. Rev. A **98**, 022112 (2018)] to witness a violation of preparation noncontextuality, that is not suitable in experiments where the operational equivalences to be tested are specified in advance, with an approach based on the notion of bounded ontological distinctness for preparations, defined by Chaturvedi and Saha [A. Chaturvedi and D. Saha, Quantum **4**, 345 (2020)]. In our approach, we test bounded ontological distinctness for two particular preparations that are relevant in certain information processing tasks in that they are associated with the even and odd parity of the bits to communicate. When there exists an ontological model where this distance is preserved we talk of *parity preservation*. Our main result provides a noise threshold under which violating parity preservation (and so bounded ontological distinctness) agrees with the established method for witnessing preparation contextuality in the simplest nontrivial scenario. This is achieved by first relating the violation of parity preservation to the quantification of contextuality in terms of inaccessible information as developed by Marvian (I. Marvian, arXiv:2003.05984.), that we also show, given the way we quantify noise, to be more robust in witnessing contextuality than Pusey's noncontextuality inequality. As an application of our findings, we treat the case of two-bit parity-oblivious multiplexing in the presence of noise. In particular, given that we have a noise threshold below which preparation contextuality holds, we use it to establish a condition for which preparation contextuality is present in the case where the probability of success exceeds that achieved by any classical strategy.

DOI: 10.1103/PhysRevA.109.032212

I. INTRODUCTION

It is important in quantum foundations to have an appropriate definition of what it means for a given quantum feature to resist an explanation within the classical worldview, i.e., an appropriate notion of nonclassicality. In this respect, one of the leading notions of nonclassicality is preparation contextuality [1], which refers to the impossibility of a theory to admit of an ontological model that represents operationally equivalent preparations as identical probability distributions on the ontic state space. A first challenge to witness such a notion in actual experiments is that the target operational equivalences to test, that are specified *a priori*, are not in general verified. A solution to this issue is to test contextuality not for these target operational equivalences, but to consider the *a posteriori* operational equivalences holding for the noisy preparations that are actually obtained in the experiment. This solution was first adopted by Mazurek *et al.* in [2]. It was then also used by Pusey in [3] to witness preparation contextuality in what is known as the *simplest nontrivial scenario*, which consists of four preparations and two binary-outcome tomographically complete measurements (Fig. 1).

Despite its success, the approach of considering the *a posteriori* operational equivalences, that we will refer to as Pusey's approach, is not suitable to treat certain information processing tasks defined in the simplest nontrivial scenario and powered by preparation contextuality, such as the two-bit parity-oblivious multiplexing (POM) [4], in the presence of noise. The reason is that in those tasks the preparations involved, and so the operational equivalences, must be specified in advance (they guarantee that the parity constraint is satisfied), and it is not possible to reanalyze the data in terms of some other sets of preparations. For this reason, in this paper we consider an alternative approach to witness nonclassicality in the simplest nontrivial scenario, and we relate it with Pusey's approach. More precisely, we consider an approach that still refers to the *a priori* ideal preparations and that is based on a more general notion of classicality named bounded ontological distinctness for preparations (BOD_P) [5].[1]

*massy@math.ucsb.edu

†lorenzo.catani4@gmail.com

[1]The simplest nontrivial scenario for testing BOD_P would consist of three rather than four preparations, as shown in [5]. Nevertheless we keep the already used terminology of Pusey in [3] that is of common use when referring to the scenario under consideration here.

 ©2024 American Physical Society

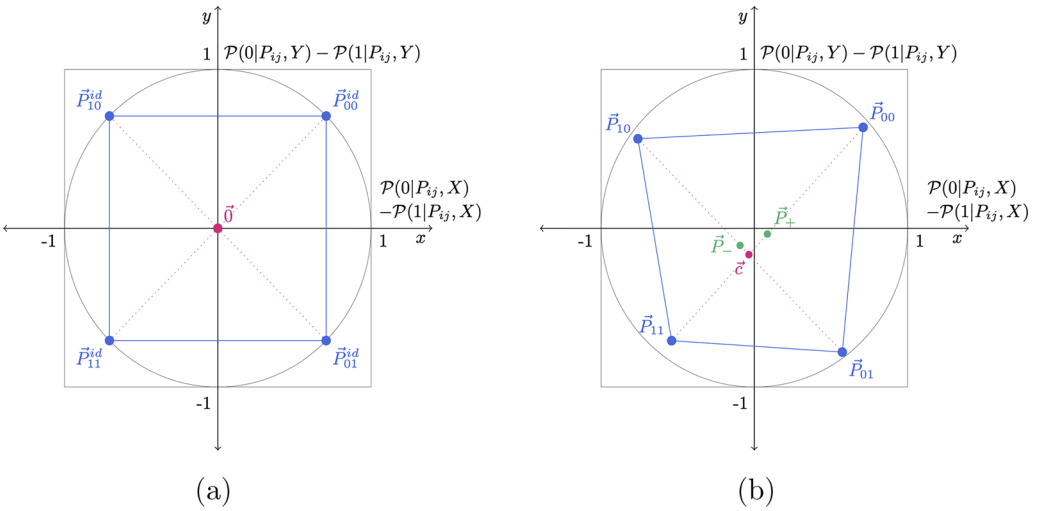

FIG. 1. The simplest nontrivial scenario. Four preparations (vertices of the blue square) and two tomographically complete measurements (corresponding to the x and y axes) are represented within the Bloch circle and the gbit square (in black). (a) The noiseless case. The *a priori* operational equivalence of Eq. (6) is represented by the vector $\vec{0}$ (in red). (b) The noisy case. The *a posteriori* operational equivalence of Eq. (7) is represented by the vector $\vec{c}$ (in red); the midpoints $\vec{P}_+, \vec{P}_-$ (in green) represent even and odd parity mixtures, respectively.

Bounded ontological distinctness for preparations generalizes the notion of preparation noncontextuality, that demands that operational equivalences are mapped to ontological equivalences, insofar as it demands that also *differences* are preserved between the operational theory and the ontological model of the theory. Unlike [5], we here define differences in terms of distances: the operational distance is the maximal gap over all measurements between the outcome probabilities of the preparations, and the ontological distance is the total variational distance between the ontic distributions [6]. With this notion of classicality, we can now treat the simplest nontrivial scenario in the presence of noise and still refer to the *a priori* ideal preparations, by addressing the distance the noisy ones have with respect to them.

We test BOD_P for the distance between the even- and odd-parity mixtures of the preparations (i.e., P_+ and P_- in Fig. 1) in the simplest nontrivial scenario. We take this to be the relevant distance because it quantifies the parity communicated in POM in the presence of noise. We refer to the case with zero difference between the operational and ontological parity distances as satisfying *parity preservation*.

In order to connect our approach based on parity preservation to Pusey's approach based on preparation noncontextuality, we use a third approach developed by Marvian in [6]. Marvian's approach quantifies preparation contextuality in terms of the "inaccessible information" of an ontological model, defined as the largest distance between a pair of ontic distributions associated to operationally equivalent preparations.

In summary, in this paper we present several results on the connection between these three approaches and on the application of such connection.

(1) Focusing on the simplest nontrivial scenario, we compare the noise thresholds for witnessing preparation contextuality in Marvian's and Pusey's approaches. We find that Marvian's is more robust,[2] in that it detects contextuality given $\delta < 0.1$, while Pusey's requires $\delta < 0.06$ (Theorems 2 and 1). Here, the noise parameter δ represents the maximum allowed deviation of the noisy experimental preparations from the corresponding noiseless target preparations in terms of maximum distinguishability with a one-shot measurement. This means that, in a one-shot measurement, the noisy preparations cannot be distinguished from the ideal preparations with probability greater than $\frac{1+\delta}{2}$.

(2) We relate preparation contextuality as witnessed in Marvian's approach and violations of parity preservation. More precisely, we provide bounds defined in terms of the operational statistics for which a violation of Marvian's equality implies a violation of parity preservation and vice versa (Theorem 3).

(3) We rewrite the above result in terms of a noise bound by taking into account the noise parameter δ (Theorem 4). As a consequence, we also find that under the threshold $\delta = 0.007$, the three approaches—Pusey's, Marvian's, and ours—agree on detecting nonclassicality; therefore an experimenter can choose the approach that they prefer in order to test nonclassicality (Theorem 6).

(4) We apply the results above to the case of POM in the presence of noise, which is played in the simplest nontrivial scenario. It has been shown that winning the game with a probability of success greater than $\frac{3}{4} + \frac{\varepsilon}{4}$ (which can be achieved with a strategy involving classical bits) implies a violation of BOD_P [5], and more precisely, of parity preservation (Theorem 8). Here ε corresponds to the parity communicated

[2]These thresholds depend on how the noise is modeled (see discussion on this in Sec. VII). The better robustness of Marvian's inequality is due to the better precision in bounding the relevant parameters in terms of the noise δ.

as a consequence of the noise. Given that the threshold for violating the parity preservation is $\delta < 0.007$, and that below this threshold preparation noncontextuality is also violated (Theorem 6), we have a condition for which POM in the presence of noise is powered by preparation contextuality, thus extending the result of the noiseless case [4]. This finding is not at all obvious, given that a violation of BOD_P does not, in general, imply a violation of preparation noncontextuality, and a probability of success greater than $\frac{3}{4} + \frac{\varepsilon}{4}$ has been proven only to imply a violation of BOD_P, but not of preparation noncontextuality. Moreover, we note that in order to obtain such a result, it is insufficient to combine the no-go theorem of the noiseless case (i.e., that a probability of success greater than $\frac{3}{4}$ implies preparation contextuality) with the fact that preparation contextuality holds whenever $\delta \leqslant 0.1$. This is because in the noisy POM one can in principle exploit the noise to communicate some parity also in classical strategies, thus improving the probability of success from $\frac{3}{4}$ to $\frac{3}{4} + \frac{\varepsilon}{4}$.

In the remainder of the paper, we first describe, in Sec. II, the simplest nontrivial scenario. In Sec. III, we provide the basic background and terminology regarding operational theories, ontological models, preparation noncontextuality, and distances. In Sec. IV, we describe Pusey's, Marvian's, and our approach based on BOD_P. In Sec. V, we connect these approaches and state our results. In Sec. VI, we treat the two-bit parity-oblivious multiplexing in the presence of noise, extending the result of the noiseless case. We conclude, in Sec. VII, by summarizing our findings and outlining future research.

II. THE SIMPLEST NONTRIVIAL SCENARIO

In this paper we focus on what is known as the simplest nontrivial scenario [3], meaning that we consider theories associated with experiments consisting of four preparations $\{P_{ij}\} = \{P_{00}, P_{01}, P_{10}, P_{11}\}$ and two binary-outcome tomographically complete measurements $\{X, Y\}$.[3] The reason why this is the *simplest* nontrivial scenario is that, as shown in Appendix B of [3], any other scenario with fewer preparations or measurements always admits the existence of a preparation noncontextual model. A geometrical representation of the scenario is depicted in Fig. 1, where the preparations are represented as vectors in the Cartesian plane with the x axis specifying the difference between the probabilities of obtaining outcomes 0 and 1 for that preparation given the measurement X, and the y axis specifying the same expression but given the measurement Y; i.e., for $i, j \in \{0, 1\}$,

$$x_{ij} = \mathcal{P}(0|P_{ij}, X) - \mathcal{P}(1|P_{ij}, X), \tag{1}$$

$$y_{ij} = \mathcal{P}(0|P_{ij}, Y) - \mathcal{P}(1|P_{ij}, Y), \tag{2}$$

$$\vec{P}_{ij} = (x_{ij}, y_{ij}). \tag{3}$$

Notice that these coordinates take values in $[-1, 1]$, and so the preparations can at maximum span the square $\vec{P}_{00} = (1, 1), \vec{P}_{01} = (-1, 1), \vec{P}_{10} = (1, -1), \vec{P}_{11} = (-1, -1)$. This square corresponds to the gbit state space [10,11], describing a nonphysical theory beyond qubit quantum theory.

With the case of quantum theory, the four preparations giving the maximum violation of noncontextuality inequalities [3,4,12,13] are (modulo the application of a unitary transformation on both preparations and measurements) denoted by $\{P^{\mathrm{id}}_{00}, P^{\mathrm{id}}_{01}, P^{\mathrm{id}}_{10}, P^{\mathrm{id}}_{11}\}$, and correspond to the vectors

$$\vec{P}^{\mathrm{id}}_{00} \equiv \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \vec{P}^{\mathrm{id}}_{01} \equiv \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right),$$
$$\vec{P}^{\mathrm{id}}_{10} \equiv \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \vec{P}^{\mathrm{id}}_{11} \equiv \left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right). \tag{4}$$

These are associated to the pure quantum states in the equator of the Bloch sphere located at angles $\frac{\pi}{4}, \frac{3\pi}{4}, \frac{5\pi}{4}, \frac{7\pi}{4}$ with respect to the $+1$ eigenstate of the Pauli X measurement. The superscript "id" stands for "ideal," stressing that these are the preparations one aims to prepare in a test of preparation contextuality in the simplest scenario.

Notably, this choice of preparations and measurements also provides the optimal quantum strategy in protocols like the two-bit parity-oblivious multiplexing [4], two-bit quantum random access codes [14], the CHSH* game [15], and several others [16].

The coordinates (operational statistics) of the vector uniquely determine the preparation it represents. It follows that two preparations P_a and P_b are operationally equivalent—denoted by $P_a \simeq P_b$—if and only if their coordinate vectors are equal:

$$P_a \simeq P_b \iff \vec{P}_a = \vec{P}_b. \tag{5}$$

Notice how the four ideal points $\vec{P}^{\mathrm{id}}_{00}, \vec{P}^{\mathrm{id}}_{01}, \vec{P}^{\mathrm{id}}_{10}$, and $\vec{P}^{\mathrm{id}}_{11}$ form two operationally equivalent decompositions of the preparation represented by the vector $\vec{0} = (0, 0)$ at the intersection of the x and y axes (the completely mixed state $\frac{I}{2}$), i.e.,

$$\tfrac{1}{2}\vec{P}^{\mathrm{id}}_{00} + \tfrac{1}{2}\vec{P}^{\mathrm{id}}_{11} = \vec{0} = \tfrac{1}{2}\vec{P}^{\mathrm{id}}_{01} + \tfrac{1}{2}\vec{P}^{\mathrm{id}}_{10}. \tag{6}$$

This operational equivalence is termed the *a priori operational equivalence*, as it is the target operational equivalence one aims to obtain in an ideal experimental test of preparation contextuality in the simplest scenario prior to performing the actual realistic experiment.

In realistic experimental scenarios, one cannot exactly prepare the ideal preparations, but necessarily obtains some noisy version of them, which we denote by $\{\vec{P}_{00}, \vec{P}_{01}, \vec{P}_{10}, \vec{P}_{11}\}$.[4] These obey an operational equivalence different from the one

[3] The assumption of tomographic completeness guarantees that there cannot exist other measurements that distinguish the operational equivalences between preparations. We refer the reader to [2,7–9] for discussions on how to experimentally deal with the assumption of tomographic completeness. Let us emphasize that X and Y are general measurements and do not represent the Pauli X and Y measurements unless we are explicitly in the case of qubit quantum theory. We adopt this notation for compatibility with the operational statistics being expressed via Cartesian coordinates.

[4] Notice that we assume that only the preparations, and not the measurements, are subjected to noise.

of Eq. (6), that we denote by the vector $\vec{c}$:

$$\underbrace{p\vec{P}_{00} + (1-p)\vec{P}_{11}}_{\vec{P}_p} = \vec{c} = \underbrace{q\vec{P}_{01} + (1-q)\vec{P}_{10}}_{\vec{P}_q} \tag{7}$$

for two probability weights $p, q \in [0, 1]$, where we denote the two operationally equivalent preparations as P_p and P_q. This latter operational equivalence is termed the *a posteriori operational equivalence*.

III. ONTOLOGICAL MODELS, NONCONTEXTUALITY, AND DISTANCES

In this section we provide the relevant background material to discuss preparation noncontextuality in the simplest scenario as well as the definitions of the distances that we use.

A. Operational theories and ontological models

The simplest scenario is an example of a prepare and measure scenario. An *operational theory* associated with a prepare and measure scenario is defined by a list of possible preparations, measurements, and the probabilities $\mathcal{P}(k|P, M)$ of obtaining the outcome k for the measurement M given that the system is prepared in the preparation P. An *ontological model* of an operational theory is meant to provide a realist explanation of the operational predictions of the theory [17]. It does so by stipulating the existence of an *ontic state space* for each given system, denoted with Λ, which is mathematically represented by a measurable set. Each point $\lambda \in \Lambda$ represents an *ontic state* that describes all the physical properties of the system. The ontological model associates each preparation P in the operational theory with a conditional probability distribution $\mu_P(\lambda) \equiv \mu(\lambda|P)$ over ontic states. We call these *epistemic states* as they represent states of knowledge about the underlying ontic states. Each measurement element $\{k, M\}$ is associated with a conditional probability distribution $\xi(k|\lambda, M)$. The latter corresponds to the probability of obtaining outcome k given that measurement M is implemented on a system in the ontic state λ. An ontological model of an operational theory reproduces the predictions of the theory via the classical law of total probability:

$$\mathcal{P}(k|P, M) = \sum_{\lambda\in\Lambda} \xi(k|\lambda, M)\mu(\lambda|P). \tag{8}$$

Here we consider ontological models for theories associated with the simplest scenario. This means theories whose preparations belong to the convex hull of the four preparations $\{P_{ij}\}$ and whose measurements are the two tomographically complete binary-outcome measurements X and Y that are assumed not to be subjected to noise.

B. Preparation noncontextuality

An ontological model is preparation noncontextual if operationally equivalent preparation procedures are represented by identical probability distributions in the ontological model [1]. More formally, two preparation procedures P and P' are operationally equivalent if they provide the same operational statistics for all possible measurements, i.e., $\forall M$: $\mathcal{P}(k|P, M) = \mathcal{P}(k|P', M)$. In this case, we write $P \simeq P'$. An ontological model is preparation noncontextual if any two such preparations are represented by the epistemic states:

$$P \simeq P' \Rightarrow \mu_P = \mu_{P'}. \tag{9}$$

An operational theory is termed *preparation noncontextual* if there exists a preparation noncontextual ontological model for the theory. Notice that, in this paper, we only consider noncontextuality with respect to preparation procedures, but the definition can be applied also to transformation and measurement procedures [1].

C. Distances

In order to quantify the distance between two probability distributions μ_a and μ_b in the ontological model we use the total variational distance, which is the standard choice (see also [6]):

$$d(\mu_a, \mu_b) \equiv \frac{1}{2}\sum_{\lambda} |\mu_a(\lambda) - \mu_b(\lambda)|. \tag{10}$$

On the other hand, we define the distances between two preparations P_a and P_b in the operational theory by the largest difference in the probabilities of measurement outcomes between two preparations across all available measurements and respective outcomes in the theory:

$$d(P_a, P_b) \equiv \max_{k,M}\{|\mathcal{P}(k|P_a, M) - \mathcal{P}(k|P_b, M)|\}. \tag{11}$$

For binary-outcome measurements, this quantity is the same for both the 0 and 1 outcome:

$$\begin{aligned} d(P_a, P_b) &= \max_{M}\{|\mathcal{P}(0|P_a, M) - \mathcal{P}(0|P_b, M)|\} \\ &= \max_{M}\{|\mathcal{P}(1|P_a, M) - \mathcal{P}(1|P_b, M)|\}. \end{aligned} \tag{12}$$

Given that there are only two measurements, X and Y, the operational distance can be equivalently expressed in the simplest scenario as follows:

$$d(P_a, P_b) \equiv \tfrac{1}{2}\max\{|x_a - x_b|, |y_a - y_b|\}. \tag{13}$$

The motivations for employing this specific operational distance are that it provides clear geometrical intuitions and that it makes the calculations tractable, as we will discuss in Sec. VII.

IV. APPROACHES

In this section we describe the three approaches considered in this paper: the first is due to Pusey [3] and provides robust noncontextuality inequalities for the simplest scenario; the second is due to Marvian [6] and witnesses preparation contextuality as the degree of what he defines as inaccessible information; the last is our approach which rephrases bounded ontological distinctness [5] in terms of the difference between the operational and ontological distances, and which specifically considers the preparations associated with the even- and odd-parity mixtures (P_+ and P_- in Fig. 1), thereby introducing the notion of parity preservation.

A. Pusey's approach

This approach allows one to test preparation contextuality in the simplest scenario considering the *a posteriori* operational equivalence of Eq. (7) and provides eight noncontextuality inequalities obtained by exploiting a connection to Bell inequalities in the famous Clauser-Horne-Shimony-Holt (CHSH) scenario [18]. The requirement of preparation noncontextuality for this *a posteriori* operational equivalence implies the existence of an ontological model for which the epistemic states μ_p and μ_q associated with the operationally equivalent preparations P_p and P_q are such that $\mu_p = \mu_q$. It is shown in [3] that this would be sufficient to make the ontological model preparation noncontextual, i.e., that any other operational equivalence within the convex hull of $\{P_{ij}\}$ corresponds to an ontological equivalence in the model.

Pusey's inequalities are denoted by $S(x_{ij}, y_{ij}) \leqslant 0$. In terms of the coordinates (x_{ij}, y_{ij}) of the noisy preparations, one can write a representative of them as

$$\begin{aligned} S(x_{ij}, y_{ij}) = {} & p(x_{00} + y_{00} + x_{11} + y_{11}) \\ & + q(x_{01} - y_{01} + x_{10} - y_{10}) \\ & + (y_{10} - x_{10} - x_{11} - y_{11}) - 2 \leqslant 0. \end{aligned} \tag{14}$$

This inequality is maximally violated, in quantum theory, by the choice of states and measurements described in the previous section [Fig. 1(a)], and the expression takes the value $S(x_{ij}^{\mathrm{id}}, y_{ij}^{\mathrm{id}}) = 4(\frac{1}{\sqrt{2}}) - 2 \approx 0.82$. The algebraic maximum value of $S(x_{ij}, y_{ij})$ is 2, which is obtained when considering the vertices of the gbit square.

We emphasize again that Pusey's approach, which considers the *a posteriori* operational equivalence, is not suitable when considering cases that require the operational equivalence to be specified in advance. An example is the case of the parity-oblivious multiplexing protocol, where the *a posteriori* operational equivalence does not embody the requirement of parity obliviousness, unlike the *a priori* operational equivalence.

B. Marvian's approach

Marvian's approach quantifies preparation contextuality through the notion of *inaccessible information*, which is defined to be the largest distance between pairs of epistemic states associated to equivalent preparation procedures, minimized over all possible ontological models:

$$C_{\mathrm{prep}}^{\mathrm{min}} \equiv \inf_{\mathrm{Models}} \sup_{P_a \simeq P_b} d(\mu_a, \mu_b). \tag{15}$$

We note that the inaccessible information $C_{\mathrm{prep}}^{\mathrm{min}}$ of an operational theory is zero if and only if the theory admits of a preparation noncontextual model:

$$C_{\mathrm{prep}}^{\mathrm{min}} = 0 \iff \text{preparation noncontextuality}. \tag{16}$$

The equality in Eq. (16) is referred to as *Marvian's preparation noncontextuality equality*.

In [6], Marvian provides a lower bound on $C_{\mathrm{prep}}^{\mathrm{min}}$ in terms of operational quantities to witness preparation contextuality. In the simplest scenario, these quantities reduce to a function of the preparations—here denoted with $\gamma(x_{ij}, y_{ij})$—the details of which are provided in Appendix B:

$$C_{\mathrm{prep}}^{\mathrm{min}} \geqslant \gamma(x_{ij}, y_{ij}). \tag{17}$$

In contrast to the notion of bounded ontological distinctness and parity preservation that are defined in the next subsections, the approaches of Marvian and Pusey both refer to the same notion: preparation noncontextuality. Therefore, they always agree in detecting contextuality in the noiseless case. However, in the noisy case and given the way we quantify noise, they provide different thresholds below which they are guaranteed to be violated. In short, we say that they provide a different *robustification*. This is why they are treated separately and it is important for our purposes to consider both.

C. BOD$_P$

BOD$_P$ was introduced in [5] and it is a criterion of classicality that requires the equivalence of the operational distinguishability between any two preparations in the operational theory and the ontological distinctness between their ontic representations, generalizing the notion of preparation noncontextuality and being based on the same credentials (a methodological principle inspired by Leibniz [19,20], or, equivalently, a principle of no operational fine tuning [21]). The *operational distinguishability* $s_{\mathcal{O}}^{P_a,P_b}$ of a pair of preparations P_a and P_b is defined as

$$s_{\mathcal{O}}^{P_a,P_b} \equiv \tfrac{1}{2} \max_{M} \{\mathcal{P}(0|P_a, M) + \mathcal{P}(1|P_b, M)\}, \tag{18}$$

where the maximum is over all measurements in the operational theory.

The *ontological distinctness* $s_{\Lambda}^{\mu_a,\mu_b}$ for the corresponding pair of epistemic states μ_a and μ_b is defined as

$$s_{\Lambda}^{\mu_a,\mu_b} \equiv \frac{1}{2} \sum_{\lambda} \max\{\mu_a(\lambda), \mu_b(\lambda)\}. \tag{19}$$

We say that an operational theory admits of a model satisfying BOD$_P$ if the value of operational distinguishability for any pair of preparations in the theory equals the value of ontological distinctness for the pair of epistemic states. That is, *bounded ontological distinctness for preparations* demands the following equality to hold for *all pairs* of preparations and corresponding epistemic states:

$$\mathrm{BOD}_P \iff s_{\mathcal{O}}^{P_a,P_b} = s_{\Lambda}^{\mu_a,\mu_b} \ \forall a, b. \tag{20}$$

To see that a model satisfying BOD$_P$ is preparation noncontextual, we consider any operationally equivalent pair P_a and P_b and observe that the expression in Eq. (18) reduces to $\frac{1}{2}$ when $P_a \simeq P_b$. Given that $s_{\mathcal{O}}^{P_a,P_b}$ represents the maximum probability of distinguishing the two preparations, the value $\frac{1}{2}$ asserts the fact that operationally equivalent procedures are completely indistinguishable. The BOD$_P$ criterion in Eq. (20) would then imply that the summation in Eq. (19) equals 1, which entails that $\mu_a(\lambda) = \mu_b(\lambda) \ \ \forall \lambda \in \Lambda$. We therefore have the implication

$$\mathrm{BOD}_P \Rightarrow \text{preparation noncontextuality}. \tag{21}$$

While the above is always true, we emphasize that the converse does not necessarily hold.

In general, an operational theory may not admit of an ontological model satisfying BOD$_P$, thus implying that there

exists a pair of preparations for which $s_\Lambda^{\mu_a,\mu_b} - s_\mathcal{O}^{P_a,P_b} > 0$. The difference $s_\Lambda^{\mu_a,\mu_b} - s_\mathcal{O}^{P_a,P_b}$ is not explicitly treated by the authors of [5], however it is of crucial relevance in the present paper given that we want to quantify this notion of nonclassicality and consider how it robustifies in the case of realistic noisy scenarios, like the simplest scenario. We quantify the violation of BOD_P as the difference of operational and ontological *distances* . Such quantification can be simply related to the original definition of BOD_P in terms of distinguishability and distinctness, as we now show. Recalling Eqs. (1), (2), and (13), and considering the available measurements in the simplest scenario, the relationship between operational distinguishability and distance is established given that

$$\begin{aligned} s_\mathcal{O}^{P_a,P_b} &= \frac{1}{2}\max_M \{\mathcal{P}(0|P_a,M) + \mathcal{P}(1|P_b,M)\} \\ &= \max\left\{\frac{1+\frac{1}{2}|x_a - x_b|}{2}, \frac{1+\frac{1}{2}|y_a - y_b|}{2}\right\} \\ &= \frac{1+d(P_a,P_b)}{2}. \end{aligned} \tag{22}$$

A similar relationship holds between the ontological distinctness and distance:

$$\begin{aligned} s_\Lambda^{\mu_a,\mu_b} &= \frac{1}{2}\sum_\lambda \max\{\mu_a(\lambda), \mu_b(\lambda)\} \\ &= \frac{1}{2}\left(1 + \frac{1}{2}\sum_\lambda |\mu_a(\lambda) - \mu_b(\lambda)|\right) \\ &= \frac{1+d(\mu_a,\mu_b)}{2}. \end{aligned} \tag{23}$$

We denote with $\mathcal{D}_{P_a,P_b}$ the difference between the ontological and operational distances for the pair of preparations P_a and P_b and their associated epistemic states. The expression $\mathcal{D}^{\min}_{P_a,P_b}$ denotes the difference $\mathcal{D}_{P_a,P_b}$ minimized over all possible ontological models. That is,

$$\mathcal{D}_{P_a,P_b} \equiv d(\mu_a,\mu_b) - d(P_a,P_b), \tag{24}$$

$$\mathcal{D}^{\min}_{P_a,P_b} \equiv \inf_{\text{Models}} \mathcal{D}_{P_a,P_b}. \tag{25}$$

Combining Eqs. (22)–(24), we view the difference of operational distinguishability and ontological distinctness as half the difference in operational and ontological distances:

$$s_\Lambda^{\mu_a,\mu_b} - s_\mathcal{O}^{P_a,P_b} = \tfrac{1}{2}\mathcal{D}_{P_a,P_b}. \tag{26}$$

Therefore $s_\Lambda^{\mu_a,\mu_b} - s_\mathcal{O}^{P_a,P_b} = 0$ if and only if $\mathcal{D}_{P_a,P_b} = 0$ and BOD_P can be equivalently expressed as the difference between operational and ontological distances being zero for all pairs:

$$\mathrm{BOD}_P \iff \mathcal{D}^{\min}_{P_a,P_b} = 0 \ \ \forall P_a, P_b. \tag{27}$$

D. Parity preservation

In this paper we test BOD_P for the difference between the operational and ontological distance, $\mathcal{D}_{P_+,P_-}$, of the even- and odd-parity mixtures, $P_+ = \frac{P_{00}+P_{11}}{2}$ and $P_- = \frac{P_{01}+P_{10}}{2}$ (see Fig. 1). The operational distance $d(P_+,P_-)$ indeed codifies the information about the parity between the bits i and j labeling the four preparations $\{P_{ij}\}$ if one measures the preparation with the measurements X and Y. For example, if $d(P_+,P_-) = 0$, then by measuring X and Y on P_+ and P_- one would always get the same outcome, thus obtaining a probability $\frac{1}{2}$ for distinguishing between them and no information about the parity between the bits i and j. If $d(P_+,P_-) \neq 0$ then some information about the parity between the bits i and j can be obtained by measuring X and Y. If $d(P_+,P_-)$ is preserved in the ontological model, meaning $\mathcal{D}_{P_+,P_-} = 0$, we say that there is *parity preservation* . Clearly, a violation of parity preservation implies a violation of BOD_P, but not vice versa. If an operational theory does not admit of a parity preserving ontological model, then $\mathcal{D}^{\min}_{P_+,P_-} > 0$. The focus on parity preservation is relevant in the context of POM and will allow us to connect a violation of BOD_P with a violation of preparation noncontextuality given certain bounds (Theorems 3 and 4).

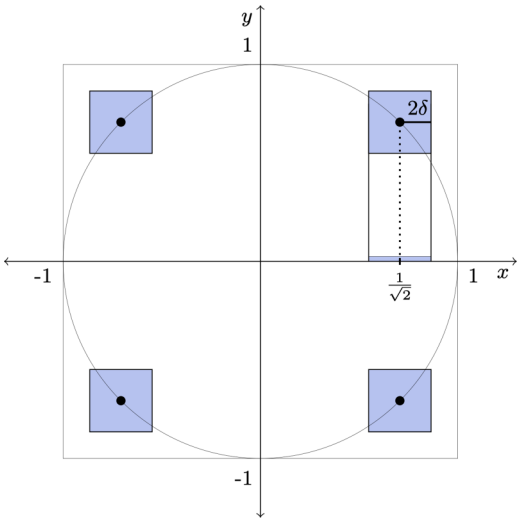

FIG. 2. Noise bound of δ. Ideal points are indicated at the center of the blue squares. These squares of radius 2δ represent regions where noisy points reside and, in accordance with Eq. (13), contain all preparations with an operational distance of *at most* δ from the ideal points, i.e., $d(\vec{P}_{ij}, \vec{P}^{\mathrm{id}}_{ij}) \leqslant \delta$. In other words, noisy points in the blue squares cannot be distinguished from their ideal counterparts with a probability greater than δ in any one-shot measurement.

V. RESULTS

In this section we report the main results (see Table I for a concise summary), all of which are proven in the Appendices. We begin with Lemma 1, which reformulates Pusey's preparation noncontextuality inequality with the specification of a noise parameter δ (see Fig. 2). This enables us to determine, in Theorem 1, a noise threshold below which the noncontextuality inequality is violated. The same is done for Marvian's preparation noncontextuality equality (Lemma 2) and a noise threshold is found below which it is still violated (Theorem 2). We continue by establishing a connection between Marvian's noncontextuality equality and parity preservation in Lemma 3. As a consequence of this lemma, in Theorem 3, we provide conditions for which a violation of one notion implies a violation of the other. That is, we show that with enough preparation contextuality, one is certain to violate

TABLE I. Three robust ways of witnessing nonclassicality in the simplest scenario. Violating Pusey's inequality and Marvian's equality are ways of witnessing preparation contextuality. Our approach is based on the notion of parity preservation. A violation of parity preservation is an instance of a violation of BOD_P. Violating parity preservation under the indicated threshold means also a violation of preparation noncontextuality.

Approach	Notion of nonclassicality	Reference to *a priori* ideal preparations	Noise threshold of violation
Pusey's	Preparation contextuality	No	$\delta < 0.06$
Marvian's	Preparation contextuality	No	$\delta < 0.1$
This paper	Violation of BOD_P	Yes	$\delta < 0.007$

parity preservation, and vice versa. The parameters for the conditions in Theorem 3 are defined from the experimental data and can be rewritten, with the aid of Lemma 4, in terms of the noise parameter δ. In this way, Theorem 3 can be rephrased into Theorem 4. The latter is used to quantify the amount of noise needed to guarantee a violation of parity preservation—Theorem 5. Finally, having obtained the noise thresholds for violating each notion of nonclassicality considered in this paper, we provide what is arguably the most relevant result—Theorem 6—that establishes a noise threshold below which all three approaches agree in witnessing nonclassicality.

A. Threshold for violating Pusey's inequality

The results of this subsection are proven in Appendix A.

Lemma 1. Suppose the preparations $\{P_{ij}\}$ of the simplest scenario satisfy a noise bound $d(\vec{P}_{ij}, \vec{P}^{\mathrm{id}}_{ij}) \leqslant \delta$, where δ is the noise parameter and $\{P^{\mathrm{id}}_{ij}\}$ are the ideal *a priori* preparations. Pusey's expression $S(x_{ij}, y_{ij})$ of Eq. (14) satisfies the following lower bound in terms of the noise parameter δ:

$$S(x_{ij}, y_{ij}) \geqslant 2\sqrt{2} - 2 - 16\delta + 32\sqrt{2}\delta^2, \tag{28}$$

where $\{x_{ij}, y_{ij}\}$ are the coordinates of the preparations $\{P_{ij}\}$.

Given that the corresponding preparation noncontextuality inequality is $S(x_{ij}, y_{ij}) \leqslant 0$, solving for the right-hand side in Eq. (28) results in a violation threshold of $\delta \approx 0.06$, which leads to the following theorem (see Fig. 3).

Theorem 1. If $d(\vec{P}_{ij}, \vec{P}^{\mathrm{id}}_{ij}) \leqslant 0.06$, then $S(x_{ij}, y_{ij}) > 0$ and Pusey's preparation noncontextuality inequality is violated.

B. Threshold for violating Marvian's equality

The results of this subsection are proven in Appendix B.

Lemma 2. Suppose the preparations $\{P_{ij}\}$ of the simplest scenario satisfy a noise bound $d(\vec{P}_{ij}, \vec{P}^{\mathrm{id}}_{ij}) \leqslant \delta$, where δ is the noise parameter and $\{P^{\mathrm{id}}_{ij}\}$ are the ideal *a priori* preparations. Marvian's inaccessible information of Eq. (15) of the scenario satisfies the following lower bound in terms of the noise parameter δ:

$$C^{\min}_{\mathrm{prep}} \geqslant \frac{\sqrt{2} - 4\delta - 1}{4(\sqrt{2} - 4\delta)}. \tag{29}$$

Given that preparation noncontextuality coincides with $C^{\min}_{\mathrm{prep}} = 0$, and we obtain a threshold of violation $\delta \approx 0.1$ when solving for the right-hand side in Eq. (29), we have the following theorem (see Fig. 3).

Theorem 2. If $d(\vec{P}_{ij}, \vec{P}^{\mathrm{id}}_{ij}) \leqslant 0.1$, then $C^{\min}_{\mathrm{prep}} > 0$ and Marvian's preparation noncontextuality equality is violated.

C. Relating Marvian's equality to parity preservation

We begin by establishing the following inequality, proven in Appendix C, connecting $\mathcal{D}^{\min}_{P_+,P_-}$ to $C^{\min}_{\mathrm{prep}}$, thus relating our approach to Marvian's.

Lemma 3. Given the simplest scenario with even- and odd-parity preparations P_+ and P_- and inaccessible information $C^{\min}_{\mathrm{prep}}$, there exist functions $\alpha_1, \alpha_2,$ and α_3 of the preparations $\{P_{ij}\}$ satisfying

$$\mathcal{D}^{\min}_{P_+,P_-} \geqslant \alpha_1 C^{\min}_{\mathrm{prep}} - \alpha_2, \tag{30a}$$

$$\mathcal{D}^{\min}_{P_+,P_-} \leqslant \alpha_1 C^{\min}_{\mathrm{prep}} + \alpha_3. \tag{30b}$$

By rearranging the terms, these relationships lead to the following theorem.

Theorem 3. Given the simplest scenario with even- and odd-parity preparations P_+ and P_- and inaccessible information $C^{\min}_{\mathrm{prep}}$, there exist functions $\alpha_1, \alpha_2,$ and α_3 of the

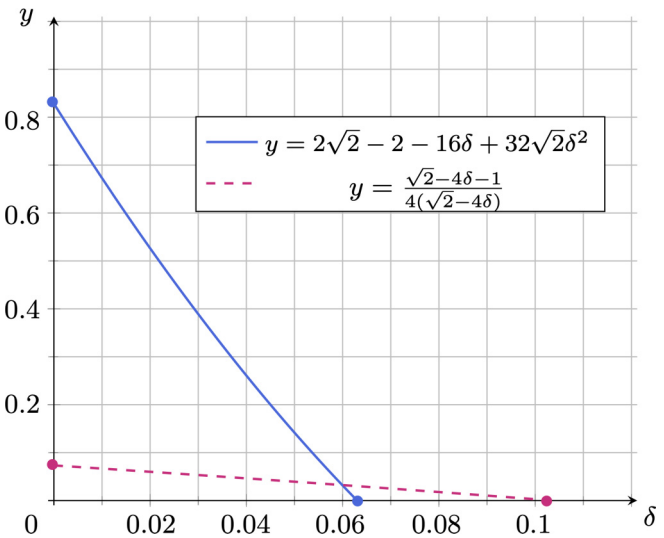

FIG. 3. Preparation contextuality witnessed by Pusey's and Marvian's approaches. The solid blue and dashed red curves plot the functions on the right-hand sides of Eqs. (28) and (29), respectively. They provide upper bounds to Pusey's and Marvian's expressions, respectively, in terms of the noise parameter δ. What is relevant in the figure is the range of δ in which each function takes a positive value, thus detecting contextuality. The specific values of the two functions are not to be compared as they are meaningful only within the scope of each approach.

preparations $\{P_{ij}\}$ such that

$$C^{\min}_{\text{prep}} > \frac{\alpha_2}{\alpha_1} \Rightarrow \mathcal{D}^{\min}_{P_+,P_-} > 0, \tag{31a}$$

$$\mathcal{D}^{\min}_{P_+,P_-} > \alpha_3 \Rightarrow C^{\min}_{\text{prep}} > 0. \tag{31b}$$

In other words, we have established that a sufficient criterion for violating parity preservation is $C^{\min}_{\text{prep}} > \frac{\alpha_2}{\alpha_1}$ and, in turn, a sufficient criterion for violating Marvian's preparation noncontextuality equality is $\mathcal{D}^{\min}_{P_+,P_-} > \alpha_3$. The parameters α_1, α_2, and α_3 are evaluated solely from the experimental data and are obtained without making reference to δ. Given any four preparations $\{P_{ij}\}$, one can directly calculate the values of α_1, α_2, and α_3 and refer to Theorem 3 to see if there is a violation of one notion of classicality given enough violation of the other. Theorem 3 can be rephrased in terms of the noise parameter δ via the following lemma, as proven in Appendix D.

Lemma 4. Given the functions α_1, α_2, and α_3, if each noisy preparation P_{ij} satisfies $d(\vec{P}_{ij}, \vec{P}^{\text{id}}_{ij}) \leqslant \delta$, the following upper bounds hold:

$$\frac{\alpha_2}{\alpha_1} \leqslant \frac{2(1+2\sqrt{3})\delta - 4\sqrt{2}\delta^2}{1-2\sqrt{2}\delta}, \tag{32a}$$

$$\alpha_3 \leqslant \frac{4\sqrt{3}\delta}{1-2\sqrt{2}\delta-4\sqrt{3}\delta}. \tag{32b}$$

Combining the previous two results, we arrive at the following statement, which recasts Theorem 3 in terms of the noise bound δ.

Theorem 4. Given the simplest scenario with even- and odd-parity preparations P_+ and P_- and inaccessible information $C^{\min}_{\text{prep}}$, if each P_{ij} satisfies $d(\vec{P}_{ij}, \vec{P}^{\text{id}}_{ij}) \leqslant \delta$, then the following implications hold:

$$C^{\min}_{\text{prep}} > \frac{2(1+2\sqrt{3})\delta - 4\sqrt{2}\delta^2}{1-2\sqrt{2}\delta} \Rightarrow \mathcal{D}^{\min}_{P_+,P_-} > 0, \tag{33a}$$

$$\mathcal{D}^{\min}_{P_+,P_-} > \frac{4\sqrt{3}\delta}{1-2\sqrt{2}\delta-4\sqrt{3}\delta} \Rightarrow C^{\min}_{\text{prep}} > 0. \tag{33b}$$

D. Threshold for nonclassicality

The results in the previous subsection ensure that for any given value of $C^{\min}_{\text{prep}}$, one can find a noise bound δ such that if noisy preparations P_{ij} lie within δ distance of the ideal preparations P^{id}_{ij}, then $\frac{\alpha_2}{\alpha_1}$ is reduced sufficiently to guarantee that indeed $C^{\min}_{\text{prep}} > \frac{\alpha_2}{\alpha_1}$ and therefore parity preservation is violated.

Equating the right-hand sides of inequalities (29) and (32a), we find the threshold for which Marvian's inequality gives a sufficient lower bound to violate parity preservation as in Eq. (31a), which results to be $\delta \approx 0.007$ (see Fig. 4).

That is, if $\delta \leqslant 0.007$, then Eq. (32a) gives us $\frac{\alpha_2}{\alpha_1} < 0.063$ whereas Eq. (29) gives us $C^{\min}_{\text{prep}} > 0.069$, so that indeed $C^{\min}_{\text{prep}} > \frac{\alpha_2}{\alpha_1}$ and $\mathcal{D}^{\min}_{P_+,P_-}$ takes on positive values. We have now established the following theorem.

Theorem 5. If $d(\vec{P}_{ij}, \vec{P}^{\text{id}}_{ij}) \leqslant 0.007$, then $\mathcal{D}^{\min}_{P_+,P_-} > 0$ and parity preservation is violated.

Given that Marvian's and Pusey's approaches both exhibit a violation if $\delta \leqslant 0.06$, we can therefore conclude sufficient conditions for equivalency in witnessing nonclassicality between all three approaches as follows.

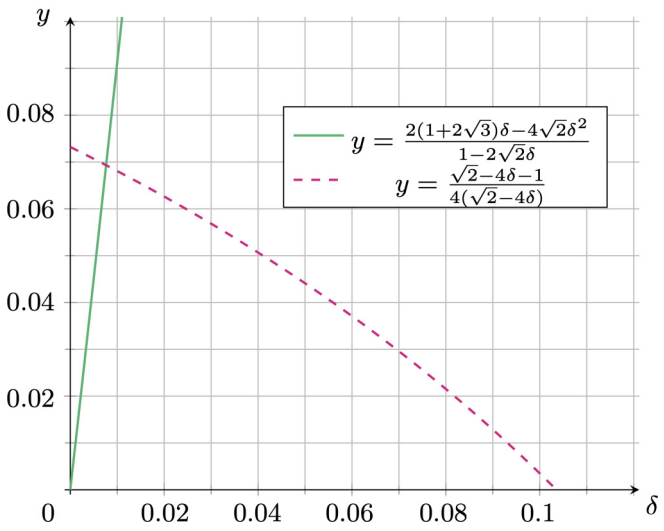

FIG. 4. Violation of parity preservation. The solid green curve indicates the function of δ in Eq. (33a). For values of δ smaller than about 0.007 (see Theorem 5), the dashed red curve upper bounding $C^{\min}_{\text{prep}}$ from Eq. (29) exceeds the solid green curve and implies, via Eq. (33a) in Theorem 4, that parity preservation is violated, i.e., $\mathcal{D}^{\min}_{P_+,P_-} > 0$.

Theorem 6. If the noise parameter δ satisfies $\delta \leqslant 0.007$, then $S(x_{ij}, y_{ij}) > 0$, $C^{\min}_{\text{prep}} > 0$, and $\mathcal{D}^{\min}_{P_+,P_-} > 0$. Therefore, all three criteria of classicality are violated.

In Appendix F, we present analogous results for the special case in which the noise manifests as quantum depolarizing noise. In this context, the noise parameter δ indicates the distance of the noisy preparation from the ideal one along the radial direction (as depicted in Fig.9). For this specific scenario, the noise threshold below which all approaches witness nonclassicality is 0.02.

VI. PARITY-OBLIVIOUS MULTIPLEXING IN THE PRESENCE OF NOISE

In this section we treat the m-bit parity-oblivious multiplexing protocol in both the noiseless and noisy scenarios. In particular, we focus on the $m = 2$ case. Indeed, in this instance, the protocol's setting involving four preparations and two measurements corresponds to the simplest scenario of Fig. 1.

A. Noiseless case

The m-bit parity-oblivious multiplexing protocol was first introduced in 2009 by Spekkens *et al.* [4]. Since then, it has motivated further investigation and substantial research on relating it to other scenarios [16,22–27], and on developing protocols with preparation contextuality as a resource for the computational advantage [28–36].

As we anticipated, we here consider the case $m = 2$. Let us imagine that Alice prepares a two-bit string, that we denote with x. Bob wishes to learn the value of a single bit among the two (without Alice knowing which one) with a probability at least p. Alice and Bob can try to achieve this by agreeing

on a strategy that consists of Alice sending some information carriers and Bob performing certain measurements. However, the task contains an additional constraint, called *parity obliviousness* : Alice cannot communicate the parity of the two-bit string x to Bob. Let us denote the bit that Bob outputs as b. The integer y denotes which of the two bits b should correspond to, and x_y denotes the actual bit in Alice's string.

The probability of success of the game takes, in general, the following form in terms of the probabilities $P(b|P_x, M_y)$, where P_x denotes the preparations of Alice, given the bit string x she has to communicate, and M_y denotes the measurements of Bob, given the bit y to be guessed:

$$\begin{aligned} p(b = x_y) = \tfrac{1}{8}[&\mathcal{P}(0|P_{00}, M_0) + \mathcal{P}(0|P_{01}, M_0) \\ &+ \mathcal{P}(1|P_{10}, M_0) + \mathcal{P}(1|P_{11}, M_0) \\ &+ \mathcal{P}(0|P_{00}, M_1) + \mathcal{P}(0|P_{10}, M_1) \\ &+ \mathcal{P}(1|P_{01}, M_1) + \mathcal{P}(1|P_{11}, M_1)]. \end{aligned} \tag{34}$$

The optimal classical probability of success satisfies $p(b = x_y) \leqslant \frac{3}{4}$, as the only classical encoding that transfers some information to Bob without violating parity obliviousness consists of encoding only a single bit x_i. Given that y is chosen at random, any bit x_i would perform the same. Therefore, Alice and Bob can agree on Alice always sending x_1 and Bob outputting $b = x_1$. The probability of success is given by the probability that $y = 1$, which is $\frac{1}{2}$, and the probability that Bob outputs correctly (at random, with probability $\frac{1}{2}$) in the other case where $y \neq 1$, that occurs with probability $\frac{1}{2}$. For this optimal classical strategy we obtain $p(b = x_y) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$, as already stated. This value is the same as the Bell bound of the CHSH game and two-bit quantum random access codes [14,18]. In [4], Spekkens *et al.* proved the following theorem (here stated only for the case $m = 2$).

Theorem 7. The optimal success probability in two-bit parity-oblivious multiplexing of any operational theory that admits of a preparation noncontextual ontological model satisfies $p(b = x_y) \leqslant \frac{3}{4}$.

This theorem indicates that preparation contextuality is a necessary resource for performing the two-bit parity-oblivious multiplexing protocol with higher success probability than the one achievable by optimal classical strategies. Moreover, it turns out that, by using the same optimal quantum strategy of two-bit quantum random access codes [14] [associated with the preparations P^{id}_{ij} and measurements X and Y of Fig. 1(a)], the probability of success is $\omega_Q(\text{POM}) = \cos^2(\frac{\pi}{8}) \approx 0.85$. It can be shown that this value is the maximum achievable with quantum strategies [4]. Notice that, when Bob measures on the X or Y basis, he cannot gain any information about the parity, as the parity 0 and parity 1 mixtures—$\frac{1}{2}P^{\text{id}}_{00} + \frac{1}{2}P^{\text{id}}_{11}$ and $\frac{1}{2}P^{\text{id}}_{01} + \frac{1}{2}P^{\text{id}}_{10}$, respectively—correspond to the same quantum state (the completely mixed state).

B. Noisy case

The two-bit parity-oblivious multiplexing protocol in realistic scenarios necessarily involves noisy preparations, that allow some parity to possibly be communicated to Bob. If one wants to generalize Theorem 7 of the noiseless case and show that preparation contextuality still powers the protocol in the noisy scenario, one faces a couple of challenges. First, one must assume that the noise could potentially be used to communicate parity, thus allowing classical strategies to achieve a probability of success greater than the ones of the noiseless case. Then, once one finds the value of the optimal classical probability of success in the noisy case, one must show that performing better than that implies a proof of preparation contextuality.

We denote the two-bit parity-oblivious multiplexing protocol in the presence of noise with $\varepsilon - \text{POM}$, where ε denotes the noise in terms of the maximum probability of parity communicated, i.e., the operational distance $d(P_+, P_-)$ between the even- and odd-parity mixtures. We first notice that, in order to witness the possible nonclassicality associated with a certain probability of success, Pusey's approach is not ideal. It considers *a posteriori* operational equivalences, thus not explicitly referring to the parity that can be communicated as a consequence of the noise and which may be part of the reason for the given probability of success. We now show how our approach based on the notion of parity preservation, that explicitly refers to the violation of parity obliviousness, is more suitable for the task.

In order to find an optimal classical strategy in $\varepsilon - \text{POM}$, we can mirror the optimal strategy described in the noiseless case of the previous subsection, this time taking into account the ε parity that can be communicated. Namely, we take into account that Bob knows the parity of x, that is, $x_1 + x_2$ (the addition is modulo 2), with probability ε. Without loss of generality, we can assume that Alice and Bob always agree on Alice sending the first bit, x_1. If $y = 1$, then Bob outputs $b = x_1$ and wins with probability 1. If $y = 2$, then with probability ε Bob knows the parity and therefore the value of x_2 [i.e., he outputs $b = x_1 + (x_1 + x_2) = x_2$ and wins with probability 1], and with probability $(1 - \varepsilon)$ he does not know the parity and can at best make a random guess about the value of x_2 and so win with probability $\frac{1}{2}$. In summary, this strategy results in the probability of success $p(b = x_y) = \frac{1}{2} \times 1 + \frac{1}{2}[\varepsilon \times 1 + (1 - \varepsilon) \times \frac{1}{2}] = \frac{3}{4} + \frac{\varepsilon}{4}$. This leads to the following lemma.[5]

Lemma 7. The optimal probability of success in two-bit ε-parity-oblivious multiplexing using a classical strategy satisfies $p(b = x_y) \leqslant \frac{3}{4} + \frac{\varepsilon}{4}$.

In Appendix E, we prove the following theorem, which is an application of Proposition 2 and the results of Sec. 4.2 of [5].

Theorem 8. The optimal success probability in two-bit parity-oblivious multiplexing of any operational theory that admits of a parity preserving ontological model satisfies $p(b = x_y) \leqslant \frac{3}{4} + \frac{\varepsilon}{4}$.

In other words, if one obtains a probability of success greater than the classical probability of success, then parity preservation is violated. Moreover, we have shown in Theorem 6 that under a certain threshold, this implies that

[5]Notice that there is no classical strategy that can perform better than the one we described here. We indeed assumed that Bob uses the knowledge of parity to his maximum possible advantage and that Alice communicates a bit of information, which is the best she can do.

Marvian's equality (as well as Pusey's inequality) is also violated. Therefore, we have shown that as long as the noise is below a certain threshold, preparation contextuality is still present whenever parity-oblivious multiplexing manifests computational advantage over classical strategies.

VII. CONCLUSION

One essential desideratum for a good notion of nonclassicality is that it should be experimentally testable. Motivated by this, we examined three approaches to test nonclassicality in the simplest nontrivial scenario, which involves four noisy preparations and two tomographically complete measurements. Specifically, we investigated Pusey's and Marvian's approaches for witnessing preparation contextuality, along with an approach for witnessing a violation of BOD_P.

We showed that these three approaches align in detecting nonclassicality as long as the level of experimental noise remains below a certain threshold, $\delta < 0.007$ (in the case of quantum depolarizing noise, this improves to $\delta < 0.02$). Therefore, experimenters have the flexibility to choose the approach that best suits their needs when testing for nonclassicality in their experiments, provided the noise remains within this range. This flexibility becomes particularly relevant in scenarios where certain approaches are not suitable, such as in the noisy parity-oblivious multiplexing protocol. In the latter case, we argued that the appropriate notion to test is parity preservation, that refers to the *a priori* ideal preparations and explicitly allows one to quantify the violation of the parity constraint. Nevertheless, by virtue of our results, below the noise threshold $\delta < 0.007$, Marvian's and Pusey's approaches can also be employed to detect nonclassicality in the experiment. Indeed, below this threshold, we also established that preparation contextuality is still present when one performs the protocol with a success probability greater than what can be achieved with classical strategies.

Crucial to obtain our results is the way we characterized noise through the noise parameter δ. The latter quantifies—via the operational distance—the deviation in the measurement statistics between the experimentally realized and the ideal target preparations. Our choice of operational distance corresponds to the maximum difference over the x and y coordinates of the statistics between the preparations. There are two reasons for employing it. First, it is geometrically intuitive, as witnessed by the fact that preparations with δ noise distance from the ideal ones belong to a square of radius 2δ around those (see Fig. 2). Second, it makes calculations tractable. We give a couple of examples.

(1) Distances over the coordinates are easily identified within Pusey's expression [Eq. (14)], making it possible to bound it.

(2) The operational distance between P_+ and P_- [Eq. (E1)] coincides with the parity that can be communicated in $\varepsilon-\mathrm{POM}$, thus allowing for a straightforward proof of Theorem 8.

With alternative definitions of operational distance like the maximum relative entropy or the total variational distance [6] we would have not exploited the above lucrative features. We leave for future research the question of how the results change by using these other ways of characterizing noise.

It is important to emphasize that determining a mathematical threshold below which both preparation noncontextuality and parity preservation are violated (Theorems 5 and 6) is not straightforward. While we expected the existence of a noise threshold below which both parity preservation (and consequently BOD_P) and preparation noncontextuality are violated, it was not clear how and if this could be found. In particular, it was not immediately obvious how to *quantitatively relate* preparation contextuality, that deals with operational equivalences and corresponding ontological inequivalences, with violations of BOD_P (and parity preservation), which deal with operational distances and corresponding greater ontological distances. The key to our achievement in obtaining such a threshold hinges on Lemma 3. The latter ultimately leads to a function of the noise parameter δ in Theorem 4 [Eq. (33a)] which intersects Marvian's witness of contextuality $C_{\mathrm{prep}}^{\mathrm{min}}$ (as shown in Fig. 4), revealing a region where not only preparation noncontextuality but also parity preservation is violated. Our success in finding this function lies in the way we recast the operational equivalence in terms of even- and odd-parity mixtures. We managed not only to allow for a connection between preparation noncontextuality and parity preservation but also to retain the amount of parity preservation violation, as shown in Appendix C. Open is the question whether a more precise threshold can be obtained, but we expect this to involve more intricate methods.

Our method based on parity preservation can be seen as an application of the results contained in [5] in the context of the simplest scenario. In connection with this previous work, it is important to highlight that we reformulated it using the concept of distances instead of distinguishabilities. This reformulation offers a clear alternative interpretation of both the operational and ontological differences. In addition, we reobtained the proof presented in [5] about the $\varepsilon-\mathrm{POM}$ being powered by a violation of BOD_P, stressing that the violation is in terms of parity preservation. Consequently, we noticed, by virtue of Theorem 6, that $\varepsilon-\mathrm{POM}$ is also powered by preparation contextuality as long as the noise parameter δ is below 0.007.

The results of this paper are relevant for applications in information processing tasks that aim to witness nonclassicality and that are set in the simplest nontrivial scenario. We recall how these tasks—examples of which are the two-bit parity-oblivious multiplexing that we treated here and other versions of the two-bit quantum random access codes—are of central importance because they are the primitive communication tasks where nonclassicality can be certified in a device or semidevice independent way. The question remains open as to whether the methods presented in this paper can be extended to scenarios beyond the simplest nontrivial case. We leave this interesting avenue for future research.

ACKNOWLEDGMENTS

The authors thank A. Chaturvedi for helpful discussions. M.K. is thankful to David Morrison for helpful discussions. This project started when L.C. was supported by the Fetzer Franklin Fund of the John E. Fetzer Memorial Trust and by ARO Grant No. W911NF-18-1-0178. L.C. also acknowledges funding from the Einstein Research Unit "Perspectives of a

Quantum Digital Transformation" and from the Digital Horizon Europe project FoQaCiA, GA Project No. 101070558. M.L. is grateful for the hospitality of Perimeter Institute where part of this work was carried out. Research at Perimeter Institute is supported in part by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Economic Development, Job Creation and Trade.

APPENDIX A: PROOF OF LEMMA 1

Suppose the preparations $\{P_{ij}\}$ of the simplest scenario satisfy a noise bound $d(\vec{P}_{ij}, \vec{P}^{\text{id}}_{ij}) \leqslant \delta$, where δ is the noise parameter and $\{P^{\text{id}}_{ij}\}$ are the ideal a priori preparations. Pusey's expression $S(x_{ij}, y_{ij})$ of Eq. (14) satisfies the following lower bound in terms of the noise parameter δ:

$$S(x_{ij}, y_{ij}) \geqslant 2\sqrt{2} - 2 - 16\delta + 32\sqrt{2}\delta^2, \tag{A1}$$

where $\{x_{ij}, y_{ij}\}$ are the coordinates of the preparations $\{P_{ij}\}$.

Proof. Let us consider Pusey's noncontextuality inequality of Eq. (14) for the case where we insert the coordinates of the preparations P_{ij} satisfying the operational equivalence (7):

$$\begin{aligned} &p(x_{00} + y_{00} + x_{11} + y_{11}) + q(x_{01} - y_{01} + x_{10} - y_{10}) \\ &\quad + (y_{10} - x_{10} - x_{11} - y_{11}) - 2 \leqslant 0. \end{aligned} \tag{A2}$$

Suppose $d(\vec{P}_{ij}, \vec{P}^{\text{id}}_{ij}) \leqslant \delta$. Given the coordinates of $\vec{P}^{\text{id}}_{ij}$ specified in Eq. (4), it follows that $l_\delta \leqslant |x_{ij}|, |y_{ij}| \leqslant u_\delta$, where $l_\delta = \frac{1}{\sqrt{2}} - 2\delta$ and $u_\delta = \frac{1}{\sqrt{2}} + 2\delta$. This can be seen in Fig. 2, where any noisy point $\vec{P}_{ij}$ within an operational distance of δ from the ideal points $\vec{P}^{\text{id}}_{ij}$ has coordinates whose absolute values lie within the range $[\frac{1}{\sqrt{2}} - 2\delta, \frac{1}{\sqrt{2}} + 2\delta]$. Further, $p, q \geqslant \frac{1-4\sqrt{2}\delta}{2}$ (see Fig. 5).

Applying these bounds to Eq. (A2), we obtain

$$\begin{aligned} S(x_{ij}, y_{ij}) &= p(x_{00} + y_{00} + x_{11} + y_{11}) \\ &\quad + q(x_{01} - y_{01} + x_{10} - y_{10}) \\ &\quad + (y_{10} - x_{10} - x_{11} - y_{11}) - 2 \\ &\geqslant \frac{1 - 4\sqrt{2}\delta}{2}(l_\delta + l_\delta - u_\delta - u_\delta) \\ &\quad + \frac{1 - 4\sqrt{2}\delta}{2}(l_\delta + l_\delta - u_\delta - u_\delta) \\ &\quad + (l_\delta + l_\delta + l_\delta + l_\delta) - 2 \\ &= 2(1 - 4\sqrt{2}\delta)(l_\delta - u_\delta) + 4l_\delta - 2 \\ &= 2(1 - 4\sqrt{2}\delta)\left(\frac{1}{\sqrt{2}} - 2\delta - \frac{1}{\sqrt{2}} - 2\delta\right) \\ &\quad + 4\left(\frac{1}{\sqrt{2}} - 2\delta\right) - 2 \\ &= 2\sqrt{2} - 2 - 16\delta + 32\sqrt{2}\delta^2. \end{aligned} \tag{A3}$$

■

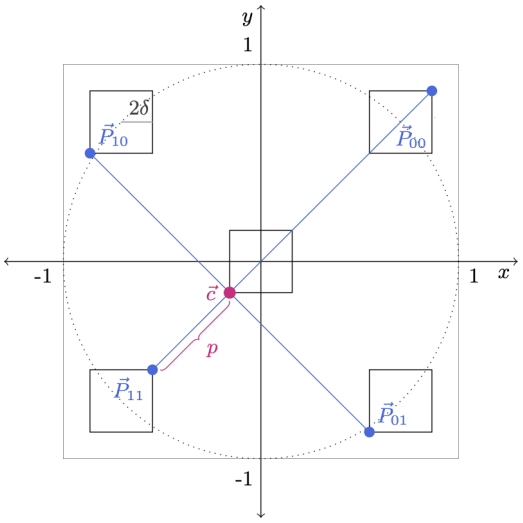

FIG. 5. An example where the weight p in Eq. (7) takes the minimum possible value, corresponding to $p = \frac{1-4\sqrt{2}\delta}{2}$. This can be seen by noting that p is the weight associated to the distance between $\vec{P}_{11}$ and $\vec{c}$ and observing that the Euclidean distance between $\vec{P}_{11}$ and $\vec{c}$ equals $1 - 4\sqrt{2}\delta$, while the Euclidean distance between $\vec{P}_{11}$ and $\vec{P}_{00}$ equals 2.

APPENDIX B: PROOF OF LEMMA 2

Suppose the preparations $\{P_{ij}\}$ of the simplest scenario satisfy a noise bound $d(\vec{P}_{ij}, \vec{P}^{\text{id}}_{ij}) \leqslant \delta$, where δ is the noise parameter and $\{P^{\text{id}}_{ij}\}$ are the ideal a priori preparations. Marvian's inaccessible information of Eq. (15) of the scenario satisfies the following lower bound in terms of the noise parameter δ:

$$C^{\min}_{\text{prep}} \geqslant \frac{\sqrt{2} - 4\delta - 1}{4(\sqrt{2} - 4\delta)}. \tag{B1}$$

Proof. We begin by using an inequality proven in [6], that provides a lower bound for the inaccessible information of an operational theory:

$$C^{\min}_{\text{prep}} \geqslant \frac{P_{\text{guess}} - \left(1 - \frac{d-1}{d}\beta^{-1}_{\min}\right)}{(d-1)d^{n-1}}. \tag{B2}$$

The values n and d represent the number of measurements in the theory and number of outcomes for each measurement, respectively. In the simplest scenario, we have $d = n = 2$. The expression P_{guess} is interpreted in [6] as the guessing probability in a certain game therein defined. In the case of the simplest scenario, $P_{\text{guess}} = 1$.[6]

[6]The preparations that give the maximal guessing probability of 1 are $\vec{R}_{00} = (1, 0)$, $\vec{R}_{01} = (-1, 0)$, $\vec{R}_{10} = (0, 1)$, and $\vec{R}_{11} = (0, -1)$. Although these four points are not in our *a priori* simplest scenario, we can augment the convex hull of the $\{\vec{P}_{ij}\}$s with the set of these additional four points, $U = \{(\pm 1, 0), (0, \pm 1)\}$, as they do not affect or contribute to the contextuality of the scenario. They indeed correspond to the stabilizer states in qubit theory, that are known not to violate preparation noncontextuality inequalities [12]. This can be also seen through Pusey's preparation noncontextuality expression, whereby (a) the calculated value of $S(x_{ij}, y_{ij})$ for U is zero and (b)

With these values, we have that Eq. (B2) reduces to the following:

$$C_{\text{prep}}^{\min} \geqslant \tfrac{1}{4}\beta_{\min}^{-1}. \tag{B3}$$

The remaining term, $\beta_{\min}$, is obtained from the operational statistics and so the right-hand side in Eq. (B3) is just a function of the preparations—it is what we referred to as $\gamma(x_{ij}, y_{ij})$ in Eq. (17). Let us now define and bound $\beta_{\min}$. Following [6], we have

$$\beta_{\min} \equiv \inf_{P} \max_{i,j} 2^{d_{\max}(P,Q_{ij})}, \tag{B4}$$

where the infimum is taken over all preparations P in the theory, $Q_{ij} \equiv \frac{1}{2}R_{0i} + \frac{1}{2}R_{1j}$ for $i, j \in \{0, 1\}$, and the distance $d_{\max}$ is the operational maximum relative entropy for preparations:

$$d_{\max}(P_a, P_b) \equiv -\log_2 \sup\{u : u \leqslant 1, \exists P_{a'} : P_b \simeq uP_a + (1-u)P_{a'}\}. \tag{B5}$$

Considering our choices of $\vec{R}_{00} = (1, 0)$, $\vec{R}_{01} = (-1, 0)$, $\vec{R}_{10} = (0, 1)$, and $\vec{R}_{11} = (0, -1)$, we have $\vec{Q}_{00} = (\frac{1}{2}, \frac{1}{2})$, $\vec{Q}_{01} = (\frac{1}{2}, -\frac{1}{2})$, $\vec{Q}_{10} = (-\frac{1}{2}, \frac{1}{2})$, and $\vec{Q}_{11} = (-\frac{1}{2}, -\frac{1}{2})$. We can now bound $\beta_{\min}$ from above:

$$\begin{aligned}
\beta_{\min} &= \inf_{P} \max_{i,j} 2^{d_{\max}(P,Q_{ij})} \\
&= \inf_{P} \max_{i,j} 2^{-\log_2 \sup\{u:u\leqslant 1, \exists P_a : Q_{ij} \simeq uP + (1-u)P_a\}} \\
&= \inf_{P} \max_{i,j} [\sup\{u : u \leqslant 1, \exists P_a : Q_{ij} \\
&\simeq uP + (1-u)P_a\}]^{-1} \\
&= \max_{i,j} \left(\sup \left\{ u : u \leqslant 1, \exists P_a : Q_{ij} \right.\right. \\
&\left.\left. \simeq u\frac{I}{2} + (1-u)P_a \right\}\right)^{-1} \\
&\leqslant \max_{i,j} \left(\left\{ u : Q_{ij} \simeq u\frac{I}{2} + (1-u)S_{ij} \right\}\right)^{-1} \\
&= \left(\left\{ u : Q_{00} \simeq u\frac{I}{2} + (1-u)S_{00} \right\}\right)^{-1} \\
&= \left(\left\{ u : \frac{1}{\sqrt{2}} = u \cdot 0 + (1-u)(1-2\sqrt{2}\delta) \right\}\right)^{-1} \\
&= \left(\frac{\sqrt{2}-4\delta-1}{\sqrt{2}-4\delta} \right)^{-1} \\
&= \frac{\sqrt{2}-4\delta}{\sqrt{2}-4\delta-1}.
\end{aligned} \tag{B6}$$

the maximal value of contextuality via $S(x_{ij}, y_{ij})$ is always attained with the initial four noisy preparations P_{ij}. We augment the set of preparations in our simplest scenario with U for the purposes of utilizing Marvian's guessing probability; the inclusion of U allows us to employ the inequality in Eq. (B2) in a manner that detects nonclassicality appropriately. That is, with $P_{\text{guess}} = 1$, the right-hand side of Eq. (B2) is positive precisely when there is preparation contextuality.

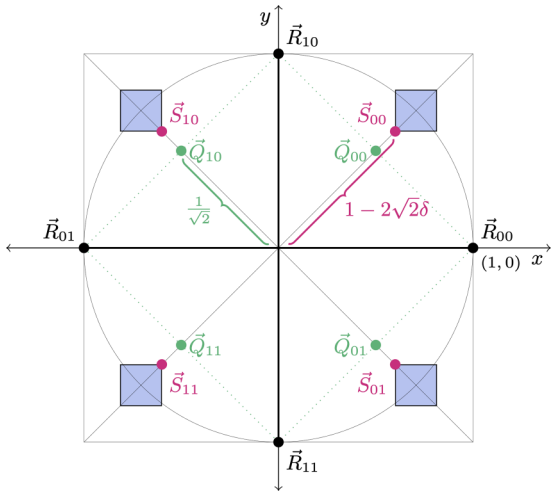

FIG. 6. The points $\vec{R}_{ij}$ correspond to the states R_{ij} used to evaluate P_{guess}. Their equal mixtures, denoted $\vec{Q}_{ij}$, are used in calculating $\beta_{\min}$. The set of $\vec{S}_{ij}$ denotes the points that are radially furthest from the $\vec{Q}_{ij}$ which are guaranteed to lie within the convex hull of the noisy preparations. That the points $\vec{S}_{ij}$ are radially furthest away from the origin (the completely mixed state) is what ensures $d_{\max}$ to be the largest distance. Note that with less noise (smaller δ) the value of $\sup\{u\}$ increases, since the stationary $\vec{Q}_{ij}$ would then be (relatively) closer to $\vec{0}$ than the $\vec{S}_{ij}$; this in turn decreases $\frac{1}{u}$, which bounds $\beta_{\min}$ from above.

In the fourth line we use the fact that, as also shown in [6], the infimum is achieved for the completely mixed state $\frac{I}{2}$. The upper bound in the fifth line arises from finding the smallest value that $\sup\{u\}$ can be guaranteed to achieve from mixing $\frac{I}{2}$ with a preparation P_a in our theory to output a fixed Q_{ij}. Any $\vec{P}_a$ that lies within the convex hull of the noisy points is a candidate to mix with $\frac{I}{2}$, and the optimal value is achieved with the preparations indicated by $\vec{S}_{ij}$ (see Fig. 6). Due to symmetry, the value of u is the same for any pair of $\vec{Q}_{ij}$ and $\vec{S}_{ij}$. In line 6 we choose, without loss of generality, $ij = 00$ to calculate the u value. We conclude with solving $\frac{1}{\sqrt{2}} = (1-u)(1-2\sqrt{2}\delta)$ for u.

Therefore $\beta_{\min}^{-1} \geqslant \frac{\sqrt{2}-4\delta-1}{\sqrt{2}-4\delta}$. This inequality applied to Eq. (B3) establishes the result. ■

APPENDIX C: PROOF OF LEMMA 3

Given the simplest scenario with even- and odd-parity preparations P_+ and P_- and inaccessible information $C_{\text{prep}}^{\min}$, there exist functions α_1, α_2, and α_3 of the preparations $\{P_{ij}\}$ satisfying

$$\mathcal{D}_{P_+,P_-}^{\min} \geqslant \alpha_1 C_{\text{prep}}^{\min} - \alpha_2, \tag{C1a}$$

$$\mathcal{D}_{P_+,P_-}^{\min} \leqslant \alpha_1 C_{\text{prep}}^{\min} + \alpha_3. \tag{C1b}$$

Proof. We first recall that

$$\underbrace{p\vec{P}_{00} + (1-p)\vec{P}_{11}}_{\vec{P}_p} = \vec{c} = \underbrace{q\vec{P}_{01} + (1-q)\vec{P}_{10}}_{\vec{P}_q}$$

and that $\vec{P}_+ = \frac{1}{2}\vec{P}_{00} + \frac{1}{2}\vec{P}_{11}$ and $\vec{P}_- = \frac{1}{2}\vec{P}_{01} + \frac{1}{2}\vec{P}_{10}$.

We then define a weight r and preparations $P_{+'}$ and $P_{-'}$ that pair with P_+ and P_- such that the following two criteria hold.

(1) $P_{+'}$ and $P_{-'}$ are *also* convex combinations of $\{P_{00}, P_{11}\}$ and $\{P_{01}, P_{10}\}$, respectively.

(2) $\vec{P}_p$ and $\vec{P}_q$ can be recast using one common weight r as

$$\underbrace{(1-r)\vec{P}_+ + r\vec{P}_{+'}}_{\vec{P}_p} = \vec{c} = \underbrace{(1-r)\vec{P}_- + r\vec{P}_{-'}}_{\vec{P}_q}. \tag{C2}$$

With Eq. (C2), we are able to write $\mu_p = (1-r)\mu_+ + r\mu_{+'}$ and $\mu_q = (1-r)\mu_- + r\mu_{-'}$. We now show how the quantities just defined allow one to lower and upper bound the distance $d(\mu_p, \mu_q)$, and so $C_{\text{prep}}^{\min}$, in terms of $d(\mu_+, \mu_-) - d(P_+, P_-)$, thus obtaining the wanted result.

We can constructively define the weight r as follows. To start, observe that $\vec{c}$ is either a convex combination of $\{\vec{P}_+, \vec{P}_{00}\}$ or $\{\vec{P}_+, \vec{P}_{11}\}$, and that the weight p determines which combination of the two pairs gives $\vec{c}$. The same is true for $\vec{c}$ being expressed as a convex combination of either $\{\vec{P}_-, \vec{P}_{01}\}$ or $\{\vec{P}_-, \vec{P}_{10}\}$, with the weight q determining which pair. With this in mind, we define the even- and odd-parity weights, r_+ and r_-, to be (here all magnitudes $||\cdot||$ indicate Euclidean length)

$$r_+ \equiv \begin{cases} \frac{||\vec{P}_+ - \vec{c}||}{||\vec{P}_+ - \vec{P}_{00}||} & p \geqslant \frac{1}{2} \\ \frac{||\vec{P}_+ - \vec{c}||}{||\vec{P}_+ - \vec{P}_{11}||} & p \leqslant \frac{1}{2} \end{cases}, \quad r_- \equiv \begin{cases} \frac{||\vec{P}_- - \vec{c}||}{||\vec{P}_- - \vec{P}_{01}||} & q \geqslant \frac{1}{2} \\ \frac{||\vec{P}_- - \vec{c}||}{||\vec{P}_- - \vec{P}_{10}||} & q \leqslant \frac{1}{2} \end{cases}. \tag{C3}$$

We now take r to be the maximum of these two values, as this ensures that our first criterion outlined earlier is satisfied, namely, that the preparations $P_{+'}$ and $P_{-'}$ can be expressed as convex combinations of $\{P_{00}, P_{11}\}$ and $\{P_{01}, P_{10}\}$, respectively. In practice, one of either $P_{+'}$ or $P_{-'}$ will equate to the original noisy P_{ij}, with the other corresponding to a strict convex combination (see Fig. 7). The specifics will depend on the weights in Eq. (7).

We now have

$$\begin{aligned} d(\mu_p, \mu_q) &= d((1-r)\mu_+ + r\mu_{+'}, (1-r)\mu_- + r\mu_{-'}) \\ &\leqslant (1-r)d(\mu_+, \mu_-) + rd(\mu_{+'}, \mu_{-'}) \\ &\leqslant (1-r)d(\mu_+, \mu_-) + r, \end{aligned} \tag{C4}$$

where the first inequality follows from the triangle inequality and the second inequality follows from $d(\mu_{+'}, \mu_{-'}) \leqslant 1$.

It therefore follows that

$$\begin{aligned} \frac{1}{1-r} d(\mu_p, \mu_q) &\leqslant d(\mu_+, \mu_-) + \frac{r}{1-r} \\ \underbrace{\frac{1}{1-r}}_{\alpha_1} d(\mu_p, \mu_q) &- \underbrace{\left[\frac{r}{(1-r)} + d(P_+, P_-)\right]}_{\alpha_2} \\ &\leqslant d(\mu_+, \mu_-) - d(P_+, P_-). \end{aligned} \tag{C5}$$

Similarly, we have

$$\begin{aligned} d(\mu_p, \mu_q) &= d((1-r)\mu_+ + r\mu_{+'}, (1-r)\mu_- + r\mu_{-'}) \\ &\geqslant (1-r)d(\mu_+, \mu_-) - rd(\mu_{+'}, \mu_{-'}) \\ &\geqslant (1-r)d(\mu_+, \mu_-) - r, \end{aligned} \tag{C6}$$

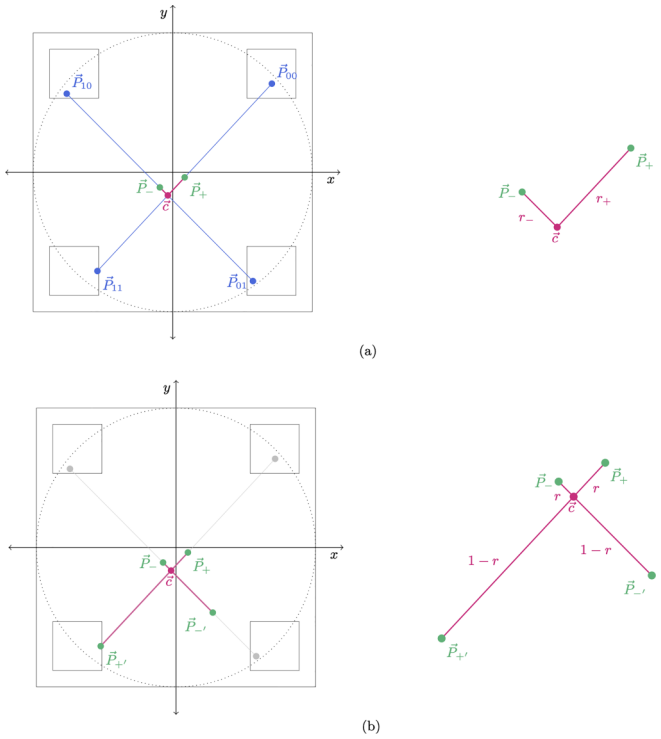

FIG. 7. (a) In this example $p < \frac{1}{2}$ and $q > \frac{1}{2}$, implying that $\vec{c}$ is a combination of $\{\vec{P}_+, \vec{P}_{11}\}$ and $\{\vec{P}_-, \vec{P}_{01}\}$, respectively, as shown on the left side. We also note that $r_+ > r_-$, as shown on the right side, so that r_+ will be used as the ratio r. (b) The intersection $\vec{c}$ has been recast (left side) in terms of two new convex combinations $\{\vec{P}_-, \vec{P}_{-'}\}$ and $\{\vec{P}_+, \vec{P}_{+'}\}$ that use the same weight r (right side). Given that these combinations are themselves mixtures of the *a priori* preparations in Eq. (7), they can be used to express $\vec{P}_p$ and $\vec{P}_q$. We note that in this example $\vec{P}_{+'} = \vec{P}_{11}$, while $\vec{P}_{-'}$ is now a strictly convex combination of $\vec{P}_-$ and $\vec{P}_{01}$. (Choosing r to have been the minimum rather than maximum would have retained $\vec{P}_{-'} = \vec{P}_{01}$; however, $\vec{P}_{+'}$ would have then been outside of the convex hull of our preparations and thus outside of the operational theory.)

which gives us

$$\begin{aligned} \frac{1}{1-r} d(\mu_p, \mu_q) &\geqslant d(\mu_+, \mu_-) - \frac{r}{1-r} \\ \underbrace{\frac{1}{1-r}}_{\alpha_1} d(\mu_p, \mu_q) &+ \underbrace{\left[\frac{r}{(1-r)} - d(P_+, P_-)\right]}_{\alpha_3} \\ &\geqslant d(\mu_+, \mu_-) - d(P_+, P_-). \end{aligned} \tag{C7}$$

In the expressions above [Eqs. (C5) and (C7)], we have identified the functions α_1, α_2, and α_3. Taking the infimum of Eqs. (C5) and (C7) over all ontological models, and noting that $C_{\text{prep}}^{\min}$ is the infimum of $d(\mu_p, \mu_q)$ across all models, the desired result is proven. ■

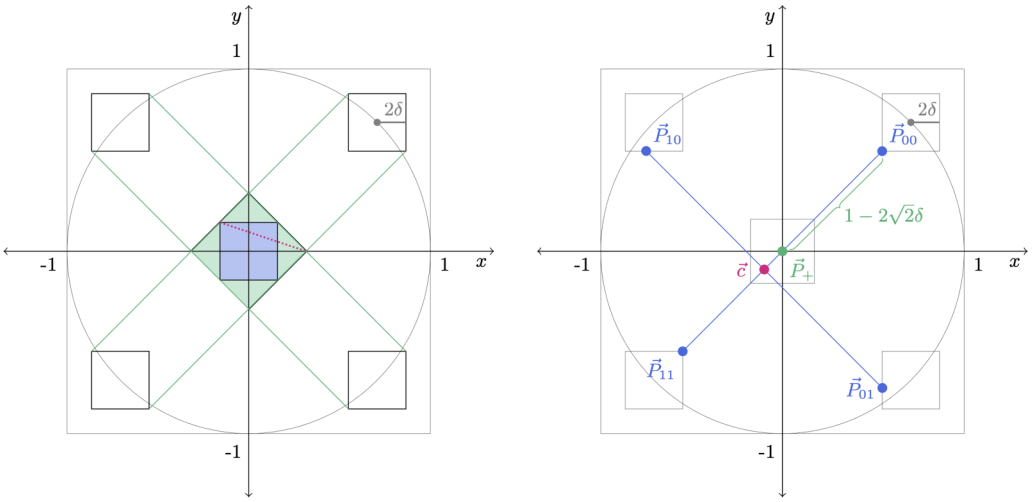

FIG. 8. (a) The point $\vec{c}$ lies within the greater rhombus region shaded in green. The points $\vec{P}_+, \vec{P}_-$ lie within the inscribed square shaded in blue. The dotted red line illustrates the maximal distance between points lying in each region, which is of length $4\sqrt{3}\delta$. (b) An example of the minimum possible Euclidean distance between a preparation vector $\vec{P}_{ij}$ and midpoint vector $\vec{P}_+$ or $\vec{P}_-$, here shown with the case of $\vec{P}_{00}$ and $\vec{P}_+$, where $||\vec{P}_+ - \vec{P}_{00}|| = 1 - 2\sqrt{2}\delta$.

APPENDIX D: PROOF OF LEMMA 4

Given the functions $\alpha_1, \alpha_2,$ andα_3, if each noisy preparation P_{ij} satisfies $d(\vec{P}_{ij}, \vec{P}_{ij}^{\text{id}}) \leqslant \delta$, the following upper bounds hold:

$$\frac{\alpha_2}{\alpha_1} \leqslant \frac{2(1+2\sqrt{3})\delta - 4\sqrt{2}\delta^2}{1-2\sqrt{2}\delta}. \tag{D1a}$$

$$\alpha_3 \leqslant \frac{\frac{4\sqrt{3}\delta}{1-2\sqrt{2}\delta}}{1-\frac{4\sqrt{3}\delta}{1-2\sqrt{2}\delta}}. \tag{D1b}$$

Proof. From Eq. (C5), we have $\alpha_1 = \frac{1}{1-r}$ and $\alpha_2 = \frac{r}{1-r} + d(P_+, P_-)$, so that

$$\frac{\alpha_2}{\alpha_1} = d(P_+, P_-)(1-r) + r. \tag{D2}$$

Note that $\vec{P}_+$ and $\vec{P}_-$ are averages of points contained in δ neighborhoods of the ideal points. Therefore, $\vec{P}_+$ and $\vec{P}_-$ are each contained in a δ neighborhood of the origin, which implies that

$$d(P_+, P_-) \leqslant 2\delta. \tag{D3}$$

Next, we observe that $\vec{c}$ is the intersection point of two line segments contained in diagonal strips of radius $2\sqrt{2}\delta$ centered in the origin. Thus $\vec{c}$ is contained within a tilted square of side length $4\sqrt{2}\delta$ centered in the origin. It follows that the maximal Euclidean distance between $\vec{P}_+$ or $\vec{P}_-$ and $\vec{c}$ is $4\sqrt{3}\delta$ [see Fig. 8(a)]. This yields

$$||\vec{P}_+ - \vec{c}||, ||\vec{P}_- - \vec{c}|| \leqslant 4\sqrt{3}\delta. \tag{D4}$$

Moreover, since $||\vec{P}_{00} - \vec{P}_+|| = \frac{1}{2}||\vec{P}_{00} - \vec{P}_{11}||$ (and similarly for the case with $\vec{P}_-$), the minimum value of the denominators in Eq. (C3) occurs when $||\vec{P}_{00} - \vec{P}_{11}||$ is minimized [see Fig. 8(b)], that corresponds to the value $2 - 4\sqrt{2}\delta$. This leads to

$$||\vec{P}_+ - \vec{P}_{00}||, ||\vec{P}_+ - \vec{P}_{11}||, ||\vec{P}_- - \vec{P}_{01}||, ||P_0 - P_{10}|| \geqslant 1 - 2\sqrt{2}\delta. \tag{D5}$$

Combining Eqs. (D4) and (D5) with Eq. (C3), it follows that

$$r = \max\{r_+, r_-\} \leqslant \frac{4\sqrt{3}\delta}{1-2\sqrt{2}\delta}. \tag{D6}$$

Referring back to Eq. (D2), this yields

$$\begin{aligned}\frac{\alpha_2}{\alpha_1} &= d(P_+, P_-)(1-r) + r \\ &\leqslant d(P_+, P_-) + r \\ &\leqslant 2\delta + \frac{4\sqrt{3}\delta}{1-2\sqrt{2}\delta} \\ &= \frac{2(1+2\sqrt{3})\delta - 4\sqrt{2}\delta^2}{1-2\sqrt{2}\delta}.\end{aligned} \tag{D7}$$

This establishes the first bound. For the second bound, we refer to Eq. (C7):

$$\begin{aligned}\alpha_3 &= \frac{r}{1-r} - d(P_+, P_-) \\ &\leqslant \frac{r}{1-r} \\ &\leqslant \frac{\frac{4\sqrt{3}\delta}{1-2\sqrt{2}\delta}}{1-\frac{4\sqrt{3}\delta}{1-2\sqrt{2}\delta}} \\ &= \frac{4\sqrt{3}\delta}{1-2\sqrt{2}\delta-4\sqrt{3}\delta}.\end{aligned} \tag{D8}$$

■

APPENDIX E: PROOF OF THEOREM 8

The optimal success probability in two-bit ε-POM of any operational theory that admits of a parity preserving ontological model satisfies $p(b = x_y) \leqslant \frac{3}{4} + \frac{\varepsilon}{4}$.

Proof. Suppose $\{P_+, P_-\}$ can be distinguished in a single-shot measurement with probability ε. That is,

$$\max_{M=X,Y} \{|\mathcal{P}(k|P_+, M) - \mathcal{P}(k|P_-, M)|\} = \varepsilon. \tag{E1}$$

This expression can be recast as $\max\{\frac{1}{2}|x_+ - x_-|, \frac{1}{2}|y_+ - y_-|\} = \varepsilon$. Therefore, by definition we have $d(P_+, P_-) = \varepsilon$. Considering Eq. (22), this means that $s_{\mathcal{O}}^{P_+,P_-} = \frac{1+\varepsilon}{2}$. Our assumption of parity preservation entails that $d(\mu_+, \mu_-) = \varepsilon$. Thus, Eq. (23) now reads as $s_{\Lambda}^{\mu_+,\mu_-} = \frac{1+\varepsilon}{2}$, which also means that $s_{\mathcal{O}}^{P_+,P_-} = s_{\Lambda}^{\mu_+,\mu_-}$. In other words, we have

$$s_{\Lambda}^{\mu_{00}+\mu_{11},\mu_{01}+\mu_{10}} = s_{\mathcal{O}}^{P_{00}+P_{11},P_{01}+P_{10}} = \frac{1+\varepsilon}{2}. \tag{E2}$$

It follows from the results of Proposition 2 in [5] that

$$\frac{1}{2}\left(s_{\mathcal{O}}^{P_{00}+P_{01},P_{10}+P_{11}}\right) + \frac{1}{2}\left(s_{\mathcal{O}}^{P_{00}+P_{10},P_{01}+P_{11}}\right) \leqslant \frac{1+\frac{1+\varepsilon}{2}}{2} = \frac{3}{4} + \frac{\varepsilon}{4}. \tag{E3}$$

The probability of success is the averaged sum of all possible ways Bob can win, given a randomly chosen measurement X, Y and state $P_{00}, P_{01}, P_{10}, P_{11}$, and an output 0,1 that Bob guesses based on the outcome of his measurement. This amounts to

$$\begin{aligned}
&p(b = x_y) \\
&= \frac{1}{8}[\mathcal{P}(0|P_{00}, X) + \mathcal{P}(0|P_{01}, X) \\
&\quad + \mathcal{P}(1|P_{10}, X) + \mathcal{P}(1|P_{11}, X) \\
&\quad + \mathcal{P}(0|P_{00}, Y) + \mathcal{P}(0|P_{10}, Y) \\
&\quad + \mathcal{P}(1|P_{01}, Y) + \mathcal{P}(1|P_{11}, Y)] \\
&= \frac{1}{2}\left(\frac{\mathcal{P}\left(0|\frac{1}{2}P_{00} + \frac{1}{2}P_{01}, X\right)}{2} + \frac{\mathcal{P}\left(1|\frac{1}{2}P_{10} + \frac{1}{2}P_{11}, X\right)}{2}\right) \\
&\quad + \frac{1}{2}\left(\frac{\mathcal{P}\left(0|\frac{1}{2}P_{00} + \frac{1}{2}P_{10}, Y\right)}{2} + \frac{\mathcal{P}\left(1|\frac{1}{2}P_{01} + \frac{1}{2}P_{11}, Y\right)}{2}\right) \\
&= \frac{1}{2}\left(s_{\mathcal{O}}^{P_{00}+P_{01},P_{10}+P_{11}}\right) + \frac{1}{2}\left(s_{\mathcal{O}}^{P_{00}+P_{10},P_{01}+P_{11}}\right).
\end{aligned} \tag{E4}$$

By combining Eqs. (E3) and (E4) we obtain $p(b = x_y) \leqslant \frac{3}{4} + \frac{\varepsilon}{4}$. ∎

APPENDIX F: THE CASE OF QUANTUM DEPOLARIZING NOISE

In this Appendix we focus on the simplest nontrivial scenario in the case where preparations are contained within the unit Bloch disk and the experimental noise is assumed to be modeled by a quantum depolarizing channel. In this case, noisy preparations P_{ij} are mixtures of the ideal preparations P_{ij}^{id} with the completely mixed state $\frac{I}{2}$ (see Fig. 9).

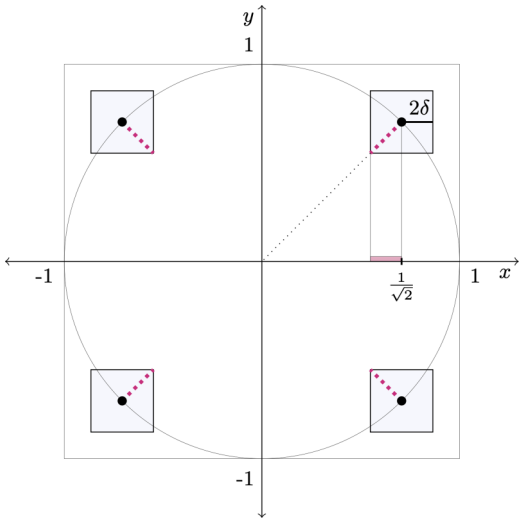

FIG. 9. Depolarizing noise constrained to a bound of δ. Ideal preparations are indicated at the center of the shaded squares. Noisy preparations are assumed to lie on the dotted red segments. In accordance with Eq. (13), they have an operational distance of *at most* δ from the ideal ones.

We begin with an updated bound on Pusey's expression and subsequent threshold for violating preparation noncontextuality based on this bound.

Lemma 8. Suppose the preparations $\{P_{ij}\}$ of the simplest scenario satisfy a noise bound $d(\vec{P}_{ij}, \vec{P}_{ij}^{\text{id}}) \leqslant \delta$, where δ is the noise parameter and $\{P_{ij}^{\text{id}}\}$ are the ideal *a priori* preparations. Pusey's expression $S(x_{ij}, y_{ij})$ of Eq. (14) satisfies the following lower bound in terms of the noise parameter δ:

$$S(x_{ij}, y_{ij}) \geqslant 2\sqrt{2} - 2 - 8\delta + \frac{8\sqrt{2}\delta^2 - 4\delta}{1 - \sqrt{2}\delta}, \tag{F1}$$

where $\{x_{ij}, y_{ij}\}$ are the coordinates of the preparations $\{P_{ij}\}$.

The proof follows that of Lemma 1 in Appendix A, with the following changes in values: $u_\delta = \frac{1}{\sqrt{2}}$ and $p, q \geqslant \frac{1-2\sqrt{2}\delta}{2-2\sqrt{2}\delta}$. Given that the corresponding preparation noncontextuality inequality is $S(x_{ij}, y_{ij}) \leqslant 0$, solving for the right-hand side in Eq. (F1) results in a violation threshold of $\delta \approx 0.07$, which leads to the following theorem.

Theorem 9. If $d(\vec{P}_{ij}, \vec{P}_{ij}^{\text{id}}) \leqslant 0.07$, then $S(x_{ij}, y_{ij}) > 0$ and Pusey's preparation noncontextuality inequality is violated.

In the case of quantum depolarizing noise, Lemma 2 and Theorem 2 are left unchanged. This is not the case for Lemma 4, that provides a different bound than the one in Eq. (32a). We provide the modified statement below.

Lemma 9. Given the functions α_1 and α_2, if each noisy preparation P_{ij} satisfies $d(\vec{P}_{ij}, \vec{P}_{ij}^{\text{id}}) \leqslant \delta$, the following upper bound holds:

$$\frac{\alpha_2}{\alpha_1} \leqslant \delta + \frac{\sqrt{2}\delta}{1 - 2\sqrt{2}\delta}. \tag{F2}$$

The proof follows that of Lemma 4 in Appendix D, with the following changes in values: $d(P_+, P_-) \leqslant \delta$; $\vec{c} = \vec{0}$; $||\vec{P}_+ - \vec{c}||, ||\vec{P}_- - \vec{c}|| \leqslant \sqrt{2}\delta$; $||\vec{P}_+ - \vec{P}_{00}||, ||\vec{P}_+ - \vec{P}_{11}||, ||\vec{P}_- - \vec{P}_{01}||, ||P_0 - P_{10}|| \geqslant 1 - 2\sqrt{2}\delta$; and $r \leqslant \frac{\sqrt{2}\delta}{1-2\sqrt{2}\delta}$.

Combining Eq. (F2) with Eq. (31a) leads to the following theorem.

Theorem 10. Given the simplest scenario with even- and odd-parity preparations P_+ and P_- and inaccessible information $C^{\min}_{\text{prep}}$, if each P_{ij} satisfies $d(\vec{P}_{ij}, \vec{P}^{\text{id}}_{ij}) \leqslant \delta$, then the following implication holds:

$$C^{\min}_{\text{prep}} > \delta + \frac{\sqrt{2}\delta}{1-2\sqrt{2}\delta} \Rightarrow \mathcal{D}^{\min}_{P_+,P_-} > 0. \tag{F3}$$

Equating the right-hand sides of inequalities (29) and (F2), we find the threshold for which Marvian's inequality provides a sufficient lower bound to violate parity preservation as in Eq. (F3), which results in $\delta \approx 0.02$. Consequently, the following theorem holds.

Theorem 11. If $d(\vec{P}_{ij}, \vec{P}^{\text{id}}_{ij}) \leqslant 0.02$, then $\mathcal{D}^{\min}_{P_+,P_-} > 0$ and parity preservation is violated.

Given that both Marvian's and Pusey's approaches exhibit a violation if $\delta \leqslant 0.07$, we obtain the following noise threshold below which all methods agree in witnessing nonclassicality in the simplest scenario in the case of quantum depolarizing noise.

Theorem 12. If the noise parameter δ satisfies $\delta \leqslant 0.02$, then $S(x_{ij}, y_{ij}) > 0$, $C^{\min}_{\text{prep}} > 0$, and $\mathcal{D}^{\min}_{P_+,P_-} > 0$. Therefore, all three criteria of classicality are violated.

[1] R. W. Spekkens, Phys. Rev. A **71**, 052108 (2005).
[2] M. D. Mazurek, M. F. Pusey, R. Kunjwal, K. J. Resch, and R. W. Spekkens, Nat. Commun. **7**, ncomms11780 (2016).
[3] M. F. Pusey, Phys. Rev. A **98**, 022112 (2018).
[4] R. W. Spekkens, D. H. Buzacott, A. J. Keehn, B. Toner, and G. J. Pryde, Phys. Rev. Lett. **102**, 010401 (2009).
[5] A. Chaturvedi and D. Saha, Quantum **4**, 345 (2020).
[6] I. Marvian, arXiv:2003.05984v1.
[7] M. F. Pusey, L. del Rio, and B. Meyer, arXiv:1904.08699.
[8] M. D. Mazurek, M. F. Pusey, K. J. Resch, and R. W. Spekkens, PRX Quantum **2**, 020302 (2021).
[9] D. Schmid, J. H. Selby, and R. W. Spekkens, arXiv:2302.07282.
[10] J. Barrett, Phys. Rev. A **75**, 032304 (2007).
[11] A. J. Short and J. Barrett, New J. Phys. **12**, 033034 (2010).
[12] L. Catani, M. Leifer, G. Scala, D. Schmid, and R. W. Spekkens, Phys. Rev. Lett. **129**, 240401 (2022).
[13] L. Catani, M. Leifer, G. Scala, D. Schmid, and R. W. Spekkens, Phys. Rev. A **108**, 022207 (2023).
[14] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani, in *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York, 1999), pp. 376–383.
[15] L. Henaut, L. Catani, D. E. Browne, S. Mansfield, and A. Pappa, Phys. Rev. A **98**, 060302(R) (2018).
[16] L. Catani, R. Faleiro, P.-E. Emeriau, S. Mansfield, and A. Pappa, Phys. Rev. A **109**, 012427 (2024).
[17] N. Harrigan and R. W. Spekkens, Found. Phys. **40**, 125 (2010).
[18] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).
[19] R. W. Spekkens, arXiv:1909.04628.
[20] D. Schmid, J. H. Selby, and R. W. Spekkens, arXiv:2009.03297.
[21] L. Catani and M. Leifer, Quantum **7**, 948 (2023).
[22] M. Banik, S. S. Bhattacharya, A. Mukherjee, A. Roy, A. Ambainis, and A. Rai, Phys. Rev. A **92**, 030103(R) (2015).
[23] A. Chailloux, I. Kerenidis, S. Kundu, and J. Sikora, New J. Phys. **18**, 045003 (2016).
[24] S. Ghorai and A. K. Pan, Phys. Rev. A **98**, 032110 (2018).
[25] D. Saha, P. Horodecki, and M. Pawłowski, New J. Phys. **21**, 093057 (2019).
[26] A. Ambainis, M. Banik, A. Chaturvedi, D. Kravchenko, and A. Rai, Quantum Info. Proc. **18**, 111 (2019).
[27] A. Tavakoli, E. Z. Cruzeiro, R. Uola, and A. A. Abbott, PRX Quantum **2**, 020334 (2021).
[28] A. Hameedi, A. Tavakoli, B. Marques, and M. Bourennane, Phys. Rev. Lett. **119**, 220402 (2017).
[29] D. Schmid and R. W. Spekkens, Phys. Rev. X **8**, 011015 (2018).
[30] D. Saha and A. Chaturvedi, Phys. Rev. A **100**, 022108 (2019).
[31] M. Lostaglio and G. Senno, Quantum **4**, 258 (2020).
[32] M. Lostaglio, Phys. Rev. Lett. **125**, 230603 (2020).
[33] S. A. Yadavalli and R. Kunjwal, Quantum **6**, 839 (2022).
[34] K. Flatt, H. Lee, C. R. I. Carceller, J. B. Brask, and J. Bae, PRX Quantum **3**, 030337 (2022).
[35] C. R. I. Carceller, K. Flatt, H. Lee, J. Bae, and J. B. Brask, Phys. Rev. Lett. **129**, 050501 (2022).
[36] R. Wagner, A. Camillini, and E. F. Galvao, Quantum **8**, 1240 (2024).