

# Verification of contextuality by noncontextual experiments

Jonathan J. Thio<sup>1</sup>, Wilfred Salmon<sup>2,3</sup>, Crispin H. W. Barnes<sup>1</sup>, Stephan De Bièvre<sup>4</sup>, and David R. M. Arvidsson-Shukur<sup>3</sup>

<sup>1</sup>*Cavendish Laboratory, Department of Physics, University of Cambridge, Cambridge CB3 0HE, United Kingdom*

<sup>2</sup>*DAMTP, Centre for Mathematical Sciences, University of Cambridge, Cambridge CB3 0WA, United Kingdom*

<sup>3</sup>*Hitachi Cambridge Laboratory, J. J. Thomson Avenue, Cambridge CB3 0HE, United Kingdom*

<sup>4</sup>*Université de Lille, CNRS, Inria, UMR 8524, Laboratoire Paul Painlevé, F-59000 Lille, France*



(Received 17 January 2025; accepted 17 November 2025; published 19 December 2025)

We uncover features of generalized contextuality by connecting it to the Kirkwood-Dirac (KD) quasiprobability distribution. Quantum states can be represented by KD distributions, which take values in the complex unit disk. Only for “KD-positive” states are the KD distributions joint probability distributions. A KD distribution can be measured by a series of weak and projective measurements. We design such an experiment and show that it is contextual iff the underlying state is not KD positive. We analyze this connection with respect to mixed KD-positive states that cannot be decomposed as convex combinations of pure KD-positive states. Our result is the construction of a noncontextual experiment that enables an experimenter to verify contextuality.

DOI: [10.1103/mzw5-t326](https://doi.org/10.1103/mzw5-t326)

## I. INTRODUCTION

Quantum physics is fundamentally different from classical physics. However, pinpointing exactly what is nonclassical is famously difficult. While certain experiments necessitate quantum theory for a full description, most can be described with classical models. This begs the question: Where does the boundary between classical and nonclassical experiments lie? Investigating this quantum-classical boundary has proven fruitful for computational [1–5] and metrological [6–10] applications. Here, we study the boundary from a more foundational perspective.

A trending and rigorous notion of nonclassicality is generalized contextuality. Although many different approaches to contextuality exist (Kochen-Specker [11–13], sheaf theoretic [14], and Contextuality-by-Default approaches [15–17], for example), we focus on generalized contextuality [18–26] which is broadly used and versatile. To determine, within this notion, if an experiment is classical, one describes it with a hidden-variable model. In such a model, one considers a set  $\Lambda$  of hidden variables. In this framework, a preparation  $P$  is described by a probability distribution  $\mu_P(\lambda)$ , where  $\lambda \in \Lambda$ . A transformation  $T$  is described by a transition matrix between hidden variables,  $\Gamma_T(\lambda'|\lambda)$ . A measurement  $M$  is described by a probability distribution on the outcome set conditioned on the hidden variable,  $\xi_M(k|\lambda)$ , where  $k$  is a specific outcome. The probability of obtaining outcome  $k$  given the 3-tuple  $(P, T, M)$  is

$$\mathbb{P}(k|P, T, M) = \int_{\Lambda \times \Lambda} d\lambda d\lambda' \mu_P(\lambda) \Gamma_T(\lambda'|\lambda) \xi_M(k|\lambda'). \quad (1)$$

An experiment is composed of a set of such 3-tuples. The hidden-variable model can then be tuned to match predictions of quantum theory or experimental data. This can always be done [18]. However, the hidden-variable model may have nonclassical features, for example, nonlocality [27]. We capture nonclassicality using the notion of generalized contextuality [18]. We say that a hidden-variable model is noncontextual iff it assigns the same probability distribution to experimentally indistinguishable procedures. We say that an experiment is contextual iff there does not exist a noncontextual hidden-variable model describing it.

Nevertheless, in this Article, we construct an experiment (Fig. 1), the contextuality of which can be verified with noncontextual procedures. Our experiment is based on recently uncovered features of the Kirkwood-Dirac (KD) quasiprobability distribution. A state  $\rho$  can be represented by a KD distribution  $Q(\rho)$ , as outlined below. For some states,  $Q(\rho)$  is a proper probability distribution. We show that, iff this is the case, our experiment admits a noncontextual hidden-variable model. Intriguingly, certain states  $\rho_*$  are described by KD probability distributions although their pure-state components cannot be. When such states are used in our experiment, one

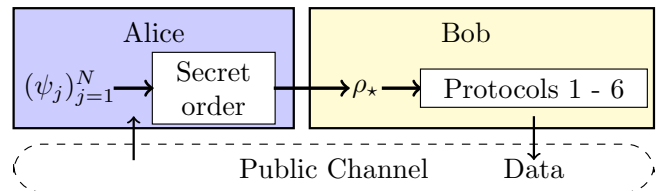


FIG. 1. Noncontextual experiment that signals contextuality. Alice sends the sequence  $(\psi_j)_{j=1}^N$  of pure states to Bob without disclosing the order. Bob effectively receives a mixed state  $\rho_* = \frac{1}{N} \sum_{j=1}^N \psi_j$ . Bob performs Procedures 1 to 6 (outlined in the main text). His experiment is noncontextual. Nonetheless, he can verify that Alice’s experiment is contextual.

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.

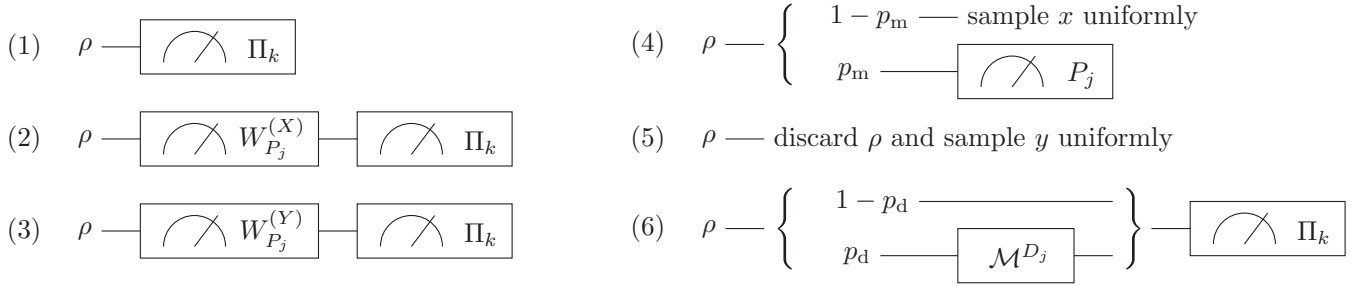


FIG. 2. Experimental protocols. Upon receiving a state  $\rho$  from Alice, Bob randomly implements one of six protocols (see Fig. 1). The boxes with  $P_j$  and  $\Pi_k$  represent projective measurements. The boxes labeled by  $W_{P_j}^{(X)}$  and  $W_{P_j}^{(Y)}$  represent  $X$ -type and  $Y$ -type weak measurements of  $P_j$ , respectively. The quantum-theoretical predictions for these protocols and their contextual implications are given in the main body of the Article.

can verify contextuality although the experiment admits a noncontextual description.

## II. KD DISTRIBUTIONS

The KD distributions [28–32] are a family of quasiprobability distributions that have recently found numerous applications in the field of quantum-information processing. We consider a Hilbert space of dimension  $d < \infty$ . The KD distribution  $Q$  maps a quantum state  $\rho$  to a corresponding quasiprobability distribution based on two nondegenerate observables  $A = \sum_j a_j P_j$  and  $B = \sum_k b_k \Pi_k$ . Here,  $\{a_j\}$  and  $\{b_k\}$  are the eigenvalues of  $A$  and  $B$ , respectively. Furthermore,  $P_j$  and  $\Pi_k$  are the rank-1 projectors onto the corresponding eigenspaces. The KD distribution of a state  $\rho$  is

$$Q_{j,k}(\rho) := \text{Tr}(\Pi_k P_j \rho). \quad (2)$$

Throughout this work, we assume that  $P_j \Pi_k \neq 0$  for all  $j$  and  $k$ . Then,  $Q$  is invertible: Knowledge of  $Q(\rho)$  enables an informationally complete reconstruction of  $\rho$  [30].

In general, the KD distribution is not a proper probability distribution as its entries may lie in the complex unit disk. However, for certain quantum states, all the entries of the KD distribution lie in the interval  $[0,1]$ , in which case the KD distribution is a probability distribution. We say that a quantum state is KD positive iff this is the case. Otherwise, we call it KD nonpositive. One can quantify how nonpositive a KD distribution is by its nonpositivity:

$$\mathcal{N}(\rho) = -1 + \sum_{j,k} |Q_{j,k}(\rho)| \geq 0, \quad (3)$$

where  $\rho$  is KD positive iff  $\mathcal{N}(\rho) = 0$  [30,33]. For example, if  $\rho$  equals an eigenstate of  $A$  or  $B$ , then  $\mathcal{N}(\rho) = 0$ . However, for pure states, if  $\rho$  has nonzero overlap with sufficiently many of the eigenstates of  $A$  and  $B$ , then  $\rho$  is KD nonpositive:  $\mathcal{N}(\rho) > 0$  [34,35]. It was recently shown, that for almost all choices of  $A$  and  $B$ , all KD-positive states are convex mixtures of  $A$ 's and  $B$ 's basis states [36]. However, for certain  $A$  and  $B$ , there exist mixed states that are KD positive but that cannot be written as convex combinations of pure KD-positive states [37]. We call such states “exotic,” and we denote the set containing them by  $\mathcal{E}_{\text{KD}^+}^{\text{exot}}$ . Exotic states will play a central role in what follows.

## III. MEASUREMENTS OF $Q(\rho)$

The experiment we design relies on the measurement of a KD distribution. To measure  $Q(\rho)$ , one can perform a series of protocols on  $\rho$ , involving the projective measurement of  $P_j$  and  $\Pi_k$ , and the so-called “weak measurement” of  $P_j$  [38–42]. The weak measurement of an observable  $A$  on a quantum state  $\rho$  is implemented as follows: (i) prepare the ancillary state  $|\psi_\epsilon\rangle := \cos \epsilon |0\rangle + \sin \epsilon |1\rangle$ , where the experimenter sets  $\epsilon \ll 1$ ; (ii) apply the interaction unitary  $U = I \otimes P_j + Z \otimes P_j^\perp$  to entangle  $\rho$  with the ancilla; and (iii) measure the ancilla in the  $X$  or  $Y$  basis. A brief calculation, given in Appendix A, shows that the Kraus operators of the  $X$ - and  $Y$ -type weak measurements of  $P_j$  are

$$N_{x,j} = \frac{1}{\sqrt{2}}(\cos \epsilon I + x \sin \epsilon D_j), \quad (4)$$

$$M_{y,j} = \frac{1}{\sqrt{2}}(\cos \epsilon I - iy \sin \epsilon D_j), \quad (5)$$

respectively. Here,  $D_j := 2P_j - I$  and  $x, y \in \{\pm 1\}$  are the outcomes of the two weak measurements. When  $\epsilon = \pi/2$ , then  $N_{x,j}$  and  $M_{x,j}$  are projective. When  $\epsilon \ll 1$ , the weak measurements convey little information about the observable.

To measure the KD distribution, we consider the following six protocols (Fig. 2) on a system in the initial state  $\rho$ . Protocols 1 through 3 measure the “weak values” [38–41] of  $P_j$ . Moreover, the addition of carefully tailored Protocols 4 through 6 enables us to uncover additional features of contextuality, as outlined below.

- (1) Measure  $\Pi_k$  on the system (returning outcome  $z$ ).
- (2) Perform an  $X$ -type weak measurement of  $P_j$  (returning outcome  $x$ ) and then measure  $\Pi_k$  (returning outcome  $z$ ).
- (3) Perform a  $Y$ -type weak measurement of  $P_j$  (returning outcome  $y$ ) and then measure  $\Pi_k$  (returning outcome  $z$ ).
- (4) Either, with probability  $p_m := \sin^2 \epsilon$ , measure  $P_j$  (returning outcome  $x$ ) or, with probability  $1 - p_m$ , return  $x \in \{\pm 1\}$  uniformly at random.
- (5) Discard the system and return as output  $y \in \{\pm 1\}$  uniformly at random.
- (6) Either apply the quantum channel  $\mathcal{M}^{D_j}(\rho) := D_j \rho D_j^\dagger$  with probability  $p_d := \sin^2 \epsilon$  or apply the identity channel with probability  $1 - p_d$ , and finally measure  $\Pi_k$  (returning outcome  $z$ ).

These six protocols include repetition over all  $j$  and  $k$ . We denote the outcome-probability distributions for these protocols by  $f_k^{(1)}(z)$ ,  $f_{j,k}^{(2)}(x, z)$ ,  $f_{j,k}^{(3)}(y, z)$ ,  $f_j^{(4)}(x)$ ,  $f_j^{(5)}(y)$ , and  $f_{j,k}^{(6)}(z)$ . Here  $x, y, z \in \{\pm 1\}$ . The quantum-theoretical predictions for the outcome-probability distributions can be calculated using Eqs. (4) and (5). We introduce the shorthand  $\Pi_k^z = \Pi_k$  if  $z = +1$  and  $\Pi_k^z = I - \Pi_k$  if  $z = -1$ . We find that

$$f_k^{(1)}(z) = \text{Tr}(\Pi_k^z \rho), \quad (6)$$

$$f_{j,k}^{(2)}(x, z) = \left(\frac{1}{2} - x\epsilon\right)\text{Tr}(\Pi_k^z \rho) + x\epsilon(1 - z)\text{Tr}(P_j \rho) + 2xz\epsilon\text{Re}(\mathcal{Q}_{j,k}(\rho)) + O(\epsilon^2), \quad (7)$$

$$f_{j,k}^{(3)}(y, z) = \frac{1}{2}\text{Tr}(\Pi_k^z \rho) + 2yz\epsilon\text{Im}\mathcal{Q}_{j,k}(\rho) + O(\epsilon^2). \quad (8)$$

Quantum theory allows us to express the outcome-probability distributions of Protocols 4 through 6 in terms of the outcome-probability distributions of Protocols 1 through 3:

$$f_j^{(4)}(x) = \sum_z f_{j,k}^{(2)}(x, z), \quad (9)$$

$$f_j^{(5)}(y) = \sum_z f_{j,k}^{(3)}(y, z), \quad (10)$$

$$f_{j,k}^{(6)}(z) = \sum_x f_{j,k}^{(2)}(x, z) = \sum_y f_{j,k}^{(3)}(y, z). \quad (11)$$

(Detailed calculations are given in Appendix A.) When  $z = +1$ , then  $x\epsilon(1 - z)\text{Tr}(P_j \rho) = 0$ , and the real and imaginary parts of the KD distribution can be deduced from Eqs. (6)–(8).

#### IV. KD NONPOSITIVITY AS A FAITHFUL WITNESS OF CONTEXTUALITY

We now establish the connection between KD nonpositivity and contextuality. Equations (9)–(11) can be used to establish noncontextuality constraints. The probability distribution for the fifth protocol can be obtained by summing over the  $z$  variable in the probability distribution for the third protocol [Eq. (10)]. This implies that, within quantum theory, one cannot distinguish (i) a  $Y$ -type weak measurement of  $P_j$  followed by a discarding of the system, and (ii) the picking of an outcome uniformly at random, entirely ignoring the system. Thus, in any noncontextual hidden-variable model, these two procedures are represented by the same probability distribution:

$$\xi_{P_j}^{\text{WY}}(y|\lambda) = \frac{1}{2}. \quad (12)$$

Here,  $\xi_{P_j}^{\text{WY}}(y|\lambda)$  is the probability distribution representing a  $Y$ -type weak measurement of  $P_j$  followed by a discarding of the system. Similarly, Eqs. (9) and (11) can be used to derive three further noncontextuality constraints:

$$\xi_{P_j}^{\text{WX}}(x|\lambda) = (1 - p_m)\frac{1}{2} + p_m\xi_{P_j}(x|\lambda), \quad (13)$$

$$\sum_{x \in \{\pm 1\}} \xi_{P_j}^{\text{WX}}(x, \lambda'|\lambda) = (1 - p_d)\delta(\lambda - \lambda') + p_d\Gamma_{D_j}(\lambda'|\lambda), \quad (14)$$

$$\sum_{y \in \{\pm 1\}} \xi_{P_j}^{\text{WY}}(y, \lambda'|\lambda) = (1 - p_d)\delta(\lambda - \lambda') + p_d\Gamma_{D_j}(\lambda'|\lambda). \quad (15)$$

Here,  $\xi_{P_j}^{\text{WX}}(x, \lambda'|\lambda)$  and  $\xi_{P_j}^{\text{WY}}(y, \lambda'|\lambda)$  are the probability distributions representing the  $X$ -type and  $Y$ -type weak measurements of  $P_j$ , respectively. Furthermore,  $\xi_{P_j}^{\text{WX}}(x|\lambda)$  is the probability distribution representing the  $X$ -type weak measurement of  $P_j$  followed by a discarding of the system;  $\xi_{P_j}(x|\lambda)$  is the probability distribution representing the measurement  $P_j$ ; and  $\Gamma_{D_j}(\lambda'|\lambda)$  is the transition matrix representing the quantum channel  $\mathcal{M}^{D_j}$ . Appendix B provides detailed derivations of these constraints.

Any noncontextual hidden-variable model for the six protocols must reproduce Eqs. (6)–(11) and satisfy Eqs. (12)–(15). For certain states  $\rho$ , such a hidden-variable model does not exist. In such a case, contextuality may be witnessed.

In the aforementioned protocols, the KD nonpositivity  $\mathcal{N}(\rho)$  is a faithful witness of contextuality:

*Theorem 1.* Assume that  $\epsilon \ll 1$  in the six protocols.

(a) If  $\rho$  is such that  $\mathcal{N}(\rho) > 3d^2\epsilon$ , then these protocols do not admit a noncontextual hidden-variable model.

(b) If  $\rho$  is KD positive [ $\mathcal{N}(\rho) = 0$ ], then these protocols admit a noncontextual hidden-variable model.

The theorem's first part is a KD rephrasing of the result proven in Refs. [43–45]. The original result is written in terms of weak values instead of the KD distribution. In Appendix C, we provide a proof along similar lines, but in terms of the KD distribution, and with an exact bound on  $\epsilon$ . This  $\epsilon$  bound will be essential to our construction of a noncontextual experiment that can verify contextuality.

The second part of the theorem has not appeared in previous literature and is essential to our results. We prove it by explicitly constructing a noncontextual hidden-variable model that has the same predictions as quantum theory [Eqs. (6)–(8)], and also obeys the Noncontextuality Constraints [Eqs. (12)–(15)]. The set of hidden variables we use is  $\Lambda = \{1, 2, \dots, d\}$ , and we represent the preparation of  $\rho$  by the outcome probability distribution when measuring the observable  $B$ :  $\mu_\rho(\lambda = j) = \text{Tr}(\rho\Pi_j)$ . In our noncontextual hidden-variable model, the other procedures are represented by the following probability distributions:

$$\xi_{\Pi_k}(z|\lambda) = \begin{cases} \delta_{k\lambda} & \text{if } z = +1 \\ 1 - \delta_{k\lambda} & \text{if } z = -1, \end{cases} \quad (16)$$

$$\begin{aligned} \xi_{P_j}^{\text{WX}}(x, \lambda'|\lambda) &= \delta_{\lambda\lambda'}\frac{1}{2}(1 - p_d) + \frac{1}{2}p_d\phi_j(\lambda'|\lambda) \\ &\quad + \frac{1}{2}x p_m \delta_{\lambda\lambda'} \left( 2\text{Re} \frac{\mathcal{Q}_{j,\lambda'}(\rho)}{\text{Tr}(\Pi_{\lambda'}\rho)} - 1 \right), \end{aligned} \quad (17)$$

$$\xi_{P_j}^{\text{WY}}(y, \lambda'|\lambda) = \delta_{\lambda\lambda'}\frac{1}{2}(1 - p_d) + \frac{1}{2}p_d\phi_j(\lambda'|\lambda). \quad (18)$$

Here,  $\phi_j(\lambda'|\lambda)$  is an arbitrary stochastic matrix. Straightforward analysis shows that these probability distributions reproduce Predictions (6)–(8). Furthermore, by substituting Eqs. (16)–(18) into the Noncontextuality Constraints (12)–(15), one can confirm that it is possible to construct  $\xi_{P_j}(x|\lambda)$  and  $\Gamma_{D_j}$  such that all noncontextuality constraints are satisfied (see Appendix D).

## V. EXPERIMENT AND ANALYSIS

We now show our main result. We construct an experiment, the contextuality of which can be verified using only noncontextual procedures. (We propose a spin-1 implementation in Appendix E.) We consider two experimenters, Alice and Bob. The result we derive concerns Bob's perspective only. Alice is introduced, for pedagogical purposes, as the experimenter that controls Bob's environment. Alice and Bob perform the following experiment:

(1) Alice chooses a sequence of  $N$  pure quantum states  $(\psi_j)_{j=1}^N$  (which may contain duplicates) such that these quantum states form an exotic state when mixed:

$$\rho_* = \frac{1}{N} \sum_{j=1}^N \psi_j. \quad (19)$$

(2) Alice sends the sequence of quantum states to Bob  $M \gg 1$  times, each time in a different order. She keeps the ordering for each sequence she sent secret.

(3) For each state Bob receives, he randomly performs one of the six protocols.

(4) Bob publicly announces which protocol he chose and what outcome he obtained for each of the  $N \times M$  states he received.

Alice's experiment involves preparing pure states, sending them to Bob, and recording the outcome that Bob announces. Bob's experiment involves taking input states prepared by Alice, and conducting one of his six protocols. Figure 1 illustrates these experiments.

Since the order of the sequence changes each time it is sent, Bob cannot distinguish individual states and thus effectively receives systems in the exotic state  $\rho_*$ . From his measurements, Bob obtains the outcome-probability distributions for each measurement procedure [Eqs. (6)–(11)] and determines  $Q(\rho_*)$ . Since  $\mathcal{N}(\rho_*) = 0$  (within some experimental error), Bob concludes, via Theorem 1, that there exists a noncontextual hidden-variable model describing his experiment.

However, knowledge of  $Q(\rho_*)$  allows for the informationally complete reconstruction of  $\rho_*$  [30]. Thus, Bob knows that  $\rho_*$  is an exotic state. In Appendix F, we show that Bob can find a  $\delta > 0$  such that any pure-state decomposition of  $\rho_*$  has at least one state  $\psi_-$  such that  $\mathcal{N}(\psi_-) > \delta$ . The outcomes of Bob's measurements are public, and Alice can analyze them for the trials where she prepared  $\psi_-$ . Bob, using the value of  $\epsilon$  he set, then checks if  $\delta > 3d^2\epsilon$ . If so, using Theorem 1, Bob deduces that there is no noncontextual hidden-variable model for Alice's postselected data. That is, Bob has verified that Alice must have performed a contextual experiment.

## VI. DISCUSSION

To summarize, we have constructed a noncontextual experiment (Bob's) in which an experimenter can verify that another experiment (Alice's) is contextual. In a sense, it is possible to verify the existence of a quantum-classical boundary from its classical side.

The nature of the exotic state  $\rho_*$  (a KD-positive state that cannot be written as a convex combination of pure KD-

positive states) is crucial to our result. We can compare the above-described scenario to one where Alice instead sends Bob states that average to the maximally mixed state  $I/d$ . Again, Bob finds his experiment to be noncontextual. However, because the maximally mixed state can be written as a convex combination of KD-positive states (for example, the eigenstates of  $A$ ), Bob cannot deduce that Alice would have verified contextuality, and thus cannot himself verify contextuality.

As mentioned, Alice may be abstracted away in our setup. Bob effectively receives a mixed state because of his limited knowledge. Nevertheless, he may infer that information exists that could “unmix” the states he has received. Thus, Bob can verify that a contextual experiment will have happened for any observer that has access to this information.

Our result has an analog in entanglement theory [46,47]. Consider a bipartite state  $\rho_{C,D}$ , spatially split between Charlie and Dave. If  $\rho_{C,D}$  is not entangled, then there exists a local hidden-variable model that describes the outcomes of any measurement conducted by Charlie and Dave [47]. The converse, however, is not true: For Werner states, the two subsystems are entangled and yet admit a local hidden-variable model [48,49]. Similarly, if Bob's experiment does not allow for the verification of quantum theory's contextuality, then it admits a noncontextual hidden-variable model. As we have shown, the converse, however, is not true: Bob's experiment can be used to verify contextuality and yet admit a noncontextual hidden-variable model. In our contextuality analysis, the exotic states take on the same role as the Werner states do in entanglement theory.

We can also compare the contextuality witness of our analysis with that of entanglement theory. We witness contextuality with KD nonpositivity  $\mathcal{N}(\rho)$ . In entanglement theory, the von Neumann entropy  $S(\rho_C) = S(\rho_D)$ , where  $\rho_C = \text{Tr}_D(\rho_{C,D})$ , etc., quantifies the entanglement of a pure state  $\rho_{C,D}$  [46,47]. The von Neumann entropy  $S(\rho_C)$  is a concave function in  $\rho_{C,D}$  equal to zero for pure separable bipartite states. For some states,  $S(\rho_C)$  exceeds zero even if  $\rho_{C,D}$  can be written as a convex combination of pure nonentangled states. Therefore,  $S(\rho_C)$  is a poor measure of mixed-state entanglement. A similar argument applies to the KD nonpositivity  $\mathcal{N}(\rho)$  with respect to contextuality.  $\mathcal{N}(\rho)$  is a convex function in  $\rho$ , equal to zero for pure KD-positive states. For some (exotic) states  $\rho_*$ ,  $\mathcal{N}(\rho_*)$  equals zero even if  $\rho_*$  cannot be written as a convex combination of pure KD-positive states. As discussed above, this means that  $\mathcal{N}(\rho)$  can be a poor witness of contextuality in our settings.

In entanglement theory, one can quantify the smallest pure-state-level entanglement of a mixed state  $\rho_{C,D}$  via the convex roof of the von Neumann entropy  $S(\rho_C)$  [47,50,51]. Similarly, one can quantify the smallest pure-state-level of KD nonpositivity of a mixed state  $\rho$  via the convex roof of  $\mathcal{N}(\rho)$  [52]. As we have seen above, such a pure-state analysis can be useful for verifying contextuality.

Finally, our result assumes quantum theory. Bob needs to reconstruct the state from his data in order to conclude that it is exotic. To do this, Bob needs to derive Eqs. (6)–(8), which requires quantum theory. Future research may strengthen our results to apply to general theories, beyond quantum theory.

### ACKNOWLEDGMENTS

J.J.T. was supported by the Cambridge Trust. W.S. was supported by the EPSRC and Hitachi. This work was supported in part by the Agence Nationale de la Recherche under Grant No. ANR-11-LABX-0007-01 (Labex CEMPI), by the Nord-Pas de Calais Regional Council and the European Regional Development Fund through the Contrat de Projets État-Région (CPER), and by the CNRS through the MITI interdisciplinary programs. S.D.B. thanks Girton College, where part of this work was performed, for its hospitality. The authors thank Christopher K. Long for his helpful comments.

### DATA AVAILABILITY

No data were created or analyzed in this study.

### APPENDIX A: THE QUANTUM THEORETICAL DESCRIPTIONS AND PREDICTIONS OF THE PROTOCOLS

In this Appendix, we describe the protocols in Fig. 2 of the Article in detail. First, we give an exact description of our weak measurement. We then calculate the quantum-theoretical predictions of the outcomes of Protocols 1 through 3. Last, we show that the outcome distributions of Protocols 4 through 6 match those obtained by marginalizing the outcome distributions of Protocols 2 and 3.

#### 1. The weak measurement

We consider a weak measurement involving a qubit pointer system [40]. The weak measurement works by entangling the system with an ancillary pointer system and then measuring the pointer in the  $X$  or the  $Y$  basis. The weak measurement thus returns a bit of information  $x \in \{\pm 1\}$  or  $y \in \{\pm 1\}$ . The strength of the weak measurement is given by the parameter  $\epsilon$ , which we take to be small. The details are given in Fig. 3.

In quantum theory, the weak measurement is represented by Kraus operators. For the  $X$ -type weak measurement given

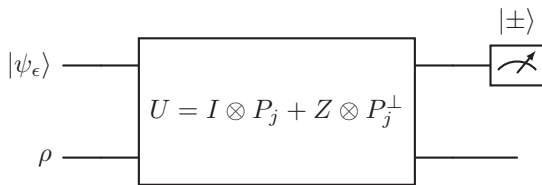


FIG. 3. Weak-Measurement Circuit. This circuit implements the  $X$ -type weak measurement on a system in an arbitrary state  $\rho$  for the observable  $P_j$ . The pointer is initialized in the state  $|\psi_\epsilon\rangle := \cos \epsilon |0\rangle + \sin \epsilon |1\rangle$ , where  $\epsilon$  controls the strength of the weak measurement. The pointer is entangled with the system via the unitary  $U = I \otimes P_j + Z \otimes P_j^\perp$ . The pointer is then measured in the  $X$  basis, giving the outcome  $x \in \{\pm 1\}$ . The circuit for the  $Y$ -type weak measurement is identical, except that the final measurement on the ancilla is replaced by a measurement in the  $Y$  basis instead of the  $X$  basis. The  $Y$ -type weak measurement returns the outcome  $y \in \{\pm 1\}$ .

in Fig. 3, the Kraus operators are

$$N_{x,j} := \langle X = x | U | \psi_\epsilon \rangle = \frac{1}{\sqrt{2}} (\cos \epsilon I + x \sin \epsilon D_j), \quad (\text{A1})$$

where  $D_j := P_j - P_j^\perp$ . Similarly, one may also calculate the Kraus operators for the  $Y$ -type weak measurement:

$$M_{y,j} := \langle Y = y | U | \psi_\epsilon \rangle = \frac{1}{\sqrt{2}} (\cos \epsilon I - iy \sin \epsilon D_j). \quad (\text{A2})$$

#### 2. Protocols 1, 2, and 3

We now use quantum theory to calculate the outcome distributions for Protocols 1 through 3 in Fig. 2 of the Article. We denote the outcome probability distributions by  $f_k^{(1)}(z)$ ,  $f_{j,k}^{(2)}(x, z)$ , and  $f_{j,k}^{(3)}(y, z)$  for each protocol, respectively.

We start by calculating the outcome probabilities for Protocol 1. This protocol involves measuring whether the system is in the  $k$ th eigenspace of observable  $B$ , and then returning  $+1$  if this is the case and  $-1$  if this is not the case. Finding the outcome  $+1$  thus corresponds to applying the projector  $\Pi_k$ , and finding the outcome  $-1$  thus corresponds to applying the projector  $\Pi_k^\perp = I - \Pi_k$ . Recall that, in the Article, we introduced the notation

$$\Pi_k^z = \begin{cases} \Pi_k, & z = +1, \\ I - \Pi_k, & z = -1. \end{cases} \quad (\text{A3})$$

The outcome probability distribution for Protocol 1 is thus

$$f_k^{(1)}(z) = \text{Tr}(\Pi_k^z \rho) := p_z^k. \quad (\text{A4})$$

We now calculate the outcome probability distribution  $f_{j,k}^{(2)}(x, z)$  for Protocol 2. The protocol involves performing an  $X$ -type weak measurement and then measuring whether or not the system is in the  $k$ th eigenspace of observable  $B$ , returning  $+1$  if this is the case and  $-1$  if it is not. The first measurement corresponds to applying the Kraus operators [see Eq. (A1)], and the second measurement corresponds to applying the same projectors as for Protocol 1. We find that

$$\begin{aligned} f_{j,k}^{(2)}(x, z) &= \mathbb{P}(x) \mathbb{P}(z|x) \\ &= \text{Tr}(N_{x,j} \rho N_{x,j}^\dagger) \text{Tr} \left[ \Pi_k^z \frac{N_{x,j} \rho N_{x,j}^\dagger}{\text{Tr}(N_{x,j} \rho N_{x,j}^\dagger)} \right] \\ &= \text{Tr}(\Pi_k^z N_{x,j} \rho N_{x,j}^\dagger) \\ &= \frac{1}{2} (1 - p_d) p_z^k + \frac{1}{2} x p_m p_z^k [2 \text{Re}(w_{j,k}^z) - 1] \\ &\quad + \frac{1}{2} p_d p_z^{D_{j,k}}. \end{aligned} \quad (\text{A5})$$

Here, we have defined the probabilities  $p_m := \sin(2\epsilon)$  and  $p_d := \sin^2(\epsilon)$ . Moreover,  $p_z^k$  is the outcome probability of measuring whether or not the system is in the  $k$ th eigenspace, and  $p_z^{D_{j,k}} := \text{Tr}(\Pi_k^z D_j \rho D_j^\dagger)$  is the outcome probability of measuring whether or not the system is in the  $k$ th eigenspace of the observable  $B$  after applying the  $D_j$  channel to the state. We have also introduced the notation

$$w_{j,k}^z = \begin{cases} w_{j,k} & \text{if } z = 1 \\ \frac{\sum_{m \neq k} w_{j,m} p_{+1}^m}{\sum_{m \neq k} p_{+1}^m} & \text{if } z = 0, \end{cases} \quad (\text{A6})$$

where  $w_{j,k}$  is the weak-value matrix [38–41] given by

$$w_{j,k} := \frac{\text{Tr}(\Pi_k P_j \rho)}{\text{Tr}(\Pi_k \rho)} = \frac{Q_{j,k}(\rho)}{\text{Tr}(\Pi_k \rho)}. \quad (\text{A7})$$

The outcome probabilities for Protocol 3 can then be calculated analogously to those for Protocol 2, except that we need to use the Kraus operators  $M_{y,j}$  defined in Eq. (A2) instead of  $N_{x,j}$ :

$$\begin{aligned} f_{j,k}^{(3)}(y, z) &= \text{Tr}(\Pi_k^z M_{y,j} \rho M_{y,j}^\dagger) \\ &= \frac{1}{2}(1 - p_d)p_z^k + y p_m p_z^k \text{Im}(w_{j,k}^z) + \frac{1}{2}p_d p_z^{D_{j,k}}. \end{aligned} \quad (\text{A8})$$

Equations (7) and (8) of the Article then follow upon approximating the right-hand side in Eqs. (A5) and (A8) to first order in  $\epsilon$  and using Eq. (A7) to write the result in terms of KD distributions.

### 3. Protocols 4, 5, and 6

We now show that, according to quantum theory, the outcome probability distributions for Protocols 4 through 6 can be obtained by marginalizing the outcome distributions of Protocols 2 and 3. Thus, if quantum theory is correct, this means that procedures (preparations, transformations, and measurements) within these protocols are experimentally indistinguishable. Consequently, we can establish noncontextuality constraints for any noncontextual hidden-variable model of the protocols.

First, we show that summation of the  $z$  variable in  $f_{j,k}^{(2)}(x, z)$  yields  $f_j^{(4)}(x)$ . Tracing over the  $z$  outcome in Protocol 2 corresponds to discarding the system after performing the  $X$ -type weak measurement. We thus effectively implement the following positive operator-valued measure (POVM) elements by tracing over  $z$ :

$$N_{x,j}^\dagger N_{x,j} = (1 - p_m)\frac{I}{2} + p_m P_j^x. \quad (\text{A9})$$

Here we introduce the notation  $P_j^x = P_j$  if  $x = +1$  and  $P_j^x = P_j^\perp$  if  $x = -1$ . Thus, we find that measuring the weak value and discarding the system is equivalent to, with probability  $1 - p_m$ , returning  $\pm 1$  uniformly at random, and with probability  $p_m$  measuring whether or not the system is in the  $j$ th eigenspace of the operator  $A$  and returning the result. Therefore, the right-hand side of Eq. (A9) exactly corresponds to the measurement process in Protocol 4:

$$\sum_z f_{j,k}^{(2)}(x, z) = f_j^{(4)}(x). \quad (\text{A10})$$

We now proceed similarly to show that the summing over  $z$  in  $f_{j,k}^{(3)}(y, z)$  yields  $f^{(5)}(y)$ . As before, summing over the  $z$  variable in Protocol 3 corresponds to the measurement process of discarding the system after performing a  $Y$ -type weak measurement. The POVM elements of this process are

$$M_y M_y^\dagger = \frac{1}{2}I. \quad (\text{A11})$$

The right-hand side of this equation corresponds to sampling  $y$  uniformly at random. This is exactly what is done in

Protocol 5. We thus find that

$$\sum_z f_{j,k}^{(3)}(y, z) = f^{(5)}(y). \quad (\text{A12})$$

Next, we show that summing over  $x$  in the outcome probability distribution  $f_{j,k}^{(2)}(x, z)$  yields the outcome probability distribution  $f_{j,k}^{(6)}(z)$ . Consider Protocol 2 without the projective measurement  $\Pi_k$ . Tracing over  $x$  corresponds to implementing the quantum channel

$$\rho \mapsto \sum_x N_{x,j} \rho N_{x,j}^\dagger = (1 - p_d)\rho + p_d D_j \rho D_j^\dagger. \quad (\text{A13})$$

$D_j$  is unitary, so this is a well-defined quantum channel. The right-hand side of Eq. (A13) is exactly the transformation process we apply in Protocol 6 before measuring  $\Pi_k$ . We thus find that

$$\sum_x f_{j,k}^{(2)}(x, z) = f_{j,k}^{(6)}(z). \quad (\text{A14})$$

A similar calculation can be used to show that

$$\sum_y f_{j,k}^{(3)}(y, z) = \sum_x f_{j,k}^{(2)}(x, z) = f_{j,k}^{(6)}(z). \quad (\text{A15})$$

## APPENDIX B: NONCONTEXTUALITY CONSTRAINTS ESTABLISHED BY PROTOCOLS 1 THROUGH 6

Here, we derive the constraints [Eqs. (12)–(15) of the Article] any noncontextual hidden-variable model of Protocols 1 through 6 must satisfy. Equation (13) of the Article was already derived in the Article. Here we will derive Eqs. (12), (14), and (15).

For clarity, in Table I, we give a list of all the procedures that are used in our six protocols. Any hidden-variable model describing those protocols must associate probability distributions to all procedures given in this table. The notation we use for every probability distribution is also outlined in this table. With a slight abuse of notation, we write  $\xi_{P_j}^{\text{WX}}(x|\lambda)$  for the probability distribution corresponding to discarding the system after an  $X$ -type weak measurement. Therefore,

$$\xi_{P_j}^{\text{WX}}(x|\lambda) = \int_\Lambda d\lambda' \xi_{P_j}^{\text{WX}}(x, \lambda'|\lambda). \quad (\text{B1})$$

We start by deriving Noncontextuality Constraint (12) in the Article, which follows from Eq. (9) in the Article:

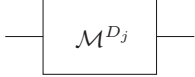
$$f_j^{(4)}(x) = \sum_z f_{j,k}^{(2)}(x, z). \quad (\text{B2})$$

This equation expresses the fact that the Protocols do not distinguish between (i) the procedure of measuring  $P_j$  returning  $x$  with probability  $p_m$  and sampling  $x$  uniformly at random with probability  $1 - p_m$ , and (ii) performing the  $X$ -type weak measurement and discarding the system afterward. Any noncontextual hidden-variable model must therefore assign the same probability distribution to these procedures:

$$\xi_{P_j}^{\text{WX}}(x|\lambda) = (1 - p_m)\frac{1}{2} + p_m \xi_{P_j}(x|\lambda). \quad (\text{B3})$$

Next, we derive Noncontextuality Constraint (14) in the Article, which follows from the first equality in Eq. (11) in

TABLE I. List of all procedures. Each row contains a description of the procedure, a pictogram representing the procedure, the corresponding mathematical object in quantum theory, and the corresponding probability distribution in the hidden-variable model.

| Description                                                                         | Pictogram                                                                         | Quantum model                                                                                       | Hidden-variable model                        |
|-------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------|----------------------------------------------|
| Prepare the state                                                                   | $\rho$ —                                                                          | Density Matrix $\rho$                                                                               | $\mu_\rho(\lambda)$                          |
| Measurement of whether the system is in the $j$ th eigenspace of the observable $A$ |  | Projectors $\{P_j, P_j^\perp = I - P_j\}$                                                           | $\xi_{P_j}(z \lambda)$                       |
| Measurement of whether the system is in the $k$ th eigenspace of the observable $B$ |  | Projectors $\{\Pi_k, \Pi_k^\perp = I - \Pi_k\}$                                                     | $\xi_{\Pi_k}(z \lambda)$                     |
| Weak measurement of $P_j$ using an $X$ -type pointer                                |  | Kraus Operators $\{N_{x,j}\}_{x \in \{\pm 1\}}$                                                     | $\xi_{P_j}^{\text{WX}}(x, \lambda' \lambda)$ |
| Weak measurement of $P_j$ using a $Y$ -type pointer                                 |  | Kraus Operators $\{M_{y,k}\}_{y \in \{\pm 1\}}$                                                     | $\xi_{P_j}^{\text{WY}}(y, \lambda' \lambda)$ |
| Quantum Channel $\mathcal{M}^{D_j}$                                                 |  | Quantum Channel $\mathcal{M}^{D_j}(\rho) = D_j \rho D_j^\dagger$ ,<br>where $D_j = P_j - P_j^\perp$ | $\Gamma_{D_j}(\lambda' \lambda)$             |

the Article:

$$f_{j,k}^{(6)}(z) = \sum_x f_{j,k}^{(2)}(x, z). \quad (\text{B4})$$

This equation expresses that the Protocols do not distinguish between (i) the transformation procedure of performing the  $X$ -type weak measurement and forgetting the result, and (ii) the transformation procedure corresponding to applying the quantum channel  $\mathcal{M}^{D_j}(\rho)$  with probability  $p_d$  and applying the identity channel with probability  $1 - p_d$ . Any noncontextual hidden-variable model must therefore assign the same probability distribution to these procedures:

$$\sum_{x \in \{\pm 1\}} \xi_{P_j}^{\text{WX}}(x, \lambda'|\lambda) = (1 - p_d)\delta(\lambda - \lambda') + p_d \Gamma_{D_j}(\lambda'|\lambda). \quad (\text{B5})$$

Last, we derive Noncontextuality Constraint (15) in the Article, which follows from the second equality in Eq. (11) in the Article:

$$f_{j,k}^{(6)}(z) = \sum_y f_{j,k}^{(3)}(y, z). \quad (\text{B6})$$

This equation implies that the Protocols do not distinguish between (i) the transformation procedure of performing the  $Y$ -type weak measurement and forgetting the result, and (ii) the transformation procedure corresponding to applying the quantum channel  $\mathcal{M}^{D_j}(\rho)$  with probability  $p_d$  and applying the identity channel with probability  $1 - p_d$ . Any noncontextual hidden-variable model must therefore assign the same probability distribution to these procedures:

$$\sum_{y \in \{\pm 1\}} \xi_{P_j}^{\text{WY}}(y, \lambda'|\lambda) = (1 - p_d)\delta(\lambda - \lambda') + p_d \Gamma_{D_j}(\lambda'|\lambda). \quad (\text{B7})$$

### APPENDIX C: KD NONPOSITIVITY IMPLIES CONTEXTUALITY

In this Appendix, we prove the first item of Theorem 1 in the Article; we show that KD nonpositivity implies contextuality for sufficiently small weak-measurement strengths  $\epsilon$ . This result is very closely related to the result in Refs. [43–45]. These previous works phrased their claims in terms of weak values and not in terms of KD distributions. Furthermore, since the previous results follow from a Taylor expansion, they do not give an explicit bound on how small  $\epsilon$  must be. We require such a bound to prove the main result of the Article.

We begin by connecting KD nonpositivity to the existence of sufficiently negative or complex elements of the KD distribution.

*Lemma C1.* Let  $\delta > 0$ . If  $\mathcal{N}(\rho) > \delta$ , then there exist indices  $j, k$  such that either  $\text{Re}Q_{j,k}(\rho) < -\frac{\delta}{3d^2}$  or  $|\text{Im}Q_{j,k}(\rho)| > \frac{\delta}{3d^2}$ .

*Proof.* We prove the contrapositive statement: if  $\rho$  is such that for all  $j, k$ ,  $\text{Re}Q_{j,k}(\rho) \geq -\frac{\delta}{3d^2}$  and  $|\text{Im}Q_{j,k}(\rho)| \leq \frac{\delta}{3d^2}$ , then  $\mathcal{N}(\rho) \leq \delta$ . Using the fact that the KD distribution is normalized, we may write the nonpositivity as

$$\mathcal{N}(\rho) = -1 + \sum_{j,k} |Q_{j,k}(\rho)| = \sum_{j,k} (|Q_{j,k}(\rho)| - \text{Re}Q_{j,k}(\rho)). \quad (\text{C1})$$

Applying the triangle inequality on the absolute value in the sum then yields

$$\mathcal{N}(\rho) \leq \sum_{j,k} (|\text{Re}Q_{j,k}(\rho)| - \text{Re}Q_{j,k}(\rho) + |\text{Im}Q_{j,k}(\rho)|). \quad (\text{C2})$$

The result then follows upon using that  $\text{Re}Q_{j,k}(\rho) \geq -\frac{\delta}{3d^2}$  and  $|\text{Im}Q_{j,k}(\rho)| \leq \frac{\delta}{3d^2}$ . ■

Next, we show that if the KD distribution is complex, then the protocols must be contextual.

**Lemma C2.** Suppose that for some state  $\rho$ , there exist  $j, k$  so that  $\text{Im}Q_{j,k}(\rho) \neq 0$ , then for  $\epsilon < \min\{|\text{Im}Q_{j,k}(\rho)|, \pi/4\}$ , Protocols 1 through 6 do not admit a noncontextual hidden-variable model.

*Proof.* We first consider the case that  $\text{Im}Q_{j,k}(\rho) > 0$ . Suppose there is a noncontextual hidden-variable model for the six protocols in Fig. 2. We will show that for sufficiently small  $\epsilon$  such a model cannot be consistent with the predictions

of quantum theory. We start by upper bounding the prediction of the hidden-variable model for  $f_3(+1, +1)$ . For  $\lambda \in \Lambda$ , let

$$I_\lambda = \{\lambda' \in \Lambda \mid \xi_{\Pi_k}(+1|\lambda') \leq \xi_{\Pi_k}(+1|\lambda)\} \quad (\text{C3})$$

be the set of ontic states with a higher “outcome +1” probability than  $\lambda$ . Using the hidden-variable model, we find that

$$f_3(+1, +1) = \int_{\Lambda} d\lambda \int_{\Lambda} d\lambda' \mu_{\rho}(\lambda) \xi_{P_j}^{\text{WY}}(+1, \lambda'|\lambda) \xi_{\Pi_k}(+1|\lambda') \quad (\text{C4})$$

$$= \int_{\Lambda} d\lambda \mu_{\rho}(\lambda) \left[ \int_{I_\lambda} d\lambda' \xi_{P_j}^{\text{WY}}(+1, \lambda'|\lambda) \xi_{\Pi_k}(+1|\lambda') + \int_{\Lambda \setminus I_\lambda} d\lambda' \xi_{P_j}^{\text{WY}}(+1, \lambda'|\lambda) \xi_{\Pi_k}(+1|\lambda') \right] \quad (\text{C5})$$

$$\leq \int_{\Lambda} d\lambda \mu_{\rho}(\lambda) \left[ \int_{I_\lambda} d\lambda' \xi_{P_j}^{\text{WY}}(+1, \lambda'|\lambda) \xi_{\Pi_k}(+1|\lambda) + \int_{\Lambda \setminus I_\lambda} d\lambda' \xi_{P_j}^{\text{WY}}(+1, \lambda'|\lambda) \right] \quad (\text{C6})$$

$$\leq \int_{\Lambda} d\lambda \mu_{\rho}(\lambda) \left[ \int_{\Lambda} d\lambda' \xi_{P_j}^{\text{WY}}(+1, \lambda'|\lambda) \xi_{\Pi_k}(+1|\lambda) + \int_{\Lambda \setminus I_\lambda} d\lambda' \sum_y \xi_{P_j}^{\text{WY}}(y, \lambda'|\lambda) \right]. \quad (\text{C7})$$

In the second line we broke up the integral; in the third line we used the definition of  $I_\lambda$  for the first term and that  $\xi_{\Pi_k}(+1|\lambda') \leq 1$  for the second term; and on the fourth line we increased the set integrated over for the first term and we summed over  $y$  in the second term.

We proceed by substituting Noncontextuality Constraints (12) and (15) from the Article into Eq. (C7). As  $\lambda \notin \Lambda \setminus I_\lambda$ , the  $\delta$  function in Eq. (15) of the Article vanishes inside the integral. We obtain

$$f_3(+1, +1) \leq \int_{\Lambda} d\lambda \mu_{\rho}(\lambda) \left[ \frac{1}{2} \xi_{\Pi_k}(+1|\lambda) + p_d \int_{\Lambda \setminus I_\lambda} d\lambda' \Gamma_{D_j}(\lambda'|\lambda) \right] \quad (\text{C8})$$

$$\leq \int_{\Lambda} d\lambda \mu_{\rho}(\lambda) \left[ \frac{1}{2} \xi_{\Pi_k}(+1|\lambda) + p_d \right] \quad (\text{C9})$$

$$= \frac{1}{2} p_{+1}^k + p_d, \quad (\text{C10})$$

where in the second line we used that integrating over any set of outcomes of a transformation matrix yields a probability less than 1.

We now turn to the quantum-theoretical prediction for  $f_3(+1, +1)$ . Equation (A8) dictates that it is

$$f_3(+1, +1) = \frac{1}{2}(1 - p_d)p_{+1}^k + p_m p_{+1}^k \text{Im}w_{j,k} + \frac{1}{2} p_z^{D_{j,k}} \quad (\text{C11})$$

$$\geq \frac{1}{2}(1 - p_d)p_{+1}^k + p_m \text{Im}Q_{j,k}(\rho). \quad (\text{C12})$$

Comparing Eqs. (C10) and (C12), we find that

$$\frac{1}{2}(1 - p_d)p_{+1}^k + p_m \text{Im}Q_{j,k}(\rho) \leq \frac{1}{2} p_{+1}^k + p_d. \quad (\text{C13})$$

Using that  $p_d = \sin^2 \epsilon$  and that  $p_m = \sin 2\epsilon$ , and rearranging, we get that

$$\tan \epsilon \geq \frac{4}{2 + p_{+1}^k} \text{Im}Q_{j,k}(\rho) \geq \frac{4}{\pi} \text{Im}Q_{j,k}(\rho), \quad (\text{C14})$$

where in the last line we used that  $2 + p_{+1}^k \leq \pi$  and that  $\text{Im}Q_{j,k}(\rho) > 0$ . For  $\epsilon \in [0, \pi/4]$ ,  $\tan \epsilon \leq \frac{4\epsilon}{\pi}$ . Thus, for  $\epsilon \in [0, \pi/4]$ , and under the assumption of a noncontextual hidden-variable model, consistent with quantum theory, we find the inequality

$$\epsilon \geq \text{Im}Q_{j,k}(\rho). \quad (\text{C15})$$

This inequality is in contradiction with the hypothesis that  $\epsilon < \min\{\text{Im}Q_{j,k}(\rho), \pi/4\}$ . We conclude that a noncontextual hidden-variable model that is consistent with quantum theory and describes Protocols 1 through 6 cannot exist for such  $\epsilon$ .

The proof for the case  $\text{Im}Q_{j,k}(\rho) < 0$  follows from considering  $f_3(-1, +1)$  and following exactly the same steps. ■

A similar result holds for the case where the KD distribution is negative:

**Lemma C3.** Suppose that for some state  $\rho$ , and some  $j, k$ ,  $\text{Re}Q_{j,k}(\rho) < 0$ ; then for  $\epsilon < \min\{-\text{Re}Q_{j,k}(\rho), \pi/4\}$ , Protocols 1 through 6 do not admit a noncontextual hidden-variable model.

*Proof.* The proof of this lemma is similar to the one of Lemma C2. We assume a noncontextual hidden-variable model exists and find a contradiction. We start by upper bounding  $f_2(-1, +1)$  using the hidden-variable model. We define  $I_\lambda$  as in the proof of Lemma C2, and through the exact same steps find that

$$f_2(-1, +1) \leq \int_{\Lambda} d\lambda \mu_{\rho}(\lambda) \left[ \int_{\Lambda} d\lambda' \xi_{P_j}^{\text{WX}}(-1, \lambda'|\lambda) \xi_{\Pi_k}(+1|\lambda) + \int_{\Lambda \setminus I_\lambda} d\lambda' \sum_x \xi_{P_j}^{\text{WX}}(x, \lambda'|\lambda) \right]. \quad (\text{C16})$$

Next, we plug Noncontextuality Constraints (13) and (14) from the Article into Eq. (C16). Using the same steps as we did for Lemma C2, we find that

$$f_2(-1, +1) \leq \frac{1}{2}(1 + p_m)p_{+1}^k + p_d. \quad (\text{C17})$$

Next, we consider the prediction by quantum theory, given by Eq. (A5):

$$f_2(-1, +1) = \frac{1}{2}(1 - p_d)p_{+1}^k - \frac{1}{2}p_m p_{+1}^k (2\text{Rew}_{j,k} - 1) + \frac{1}{2}p_d p_{+1}^{D_{j,k}} \quad (\text{C18})$$

$$\geq \frac{1}{2}(1 - p_d)p_{+1}^k - p_m \text{Re}Q_{j,k}(\rho) + \frac{1}{2}p_m p_{+1}^k. \quad (\text{C19})$$

We now combine Eqs. (C17) and (C19) and use that  $p_d = \sin^2 \epsilon$  and that  $p_m = \sin 2\epsilon$  to get that

$$\tan \epsilon \geq -\frac{4}{2 + p_{+1}^k} \text{Re}Q_{j,k}(\rho) \geq -\frac{4}{\pi} Q_{j,k}(\rho). \quad (\text{C20})$$

Again, for  $\epsilon \in [0, \pi/4]$ ,  $\tan \epsilon \leq \frac{\pi}{4}\epsilon$ , so we get that the existence of a noncontextual hidden-variable model implies that

$$\epsilon \geq -\text{Re}Q_{j,k}(\rho), \quad (\text{C21})$$

from which the Lemma follows. ■

Combining these lemmas, we find the following:

*Theorem C1.* Let  $\delta \in (0, \pi/4]$ . If  $\mathcal{N}(\rho) > \delta$ , then, for all  $\epsilon < \frac{\delta}{3\sqrt{2}}$ , Protocols 1 through 6 do not admit a noncontextual hidden-variable model.

#### APPENDIX D: KD POSITIVITY IMPLIES NONCONTEXTUALITY

In this Appendix, we prove the second item in Theorem 1 of the Article: for sufficiently small  $\epsilon$ , KD positivity implies that there exists a noncontextual hidden-variable model. Let us formally restate what we are proving here:

*Lemma D1.* If  $\rho$  is KD positive and  $\epsilon < \frac{\sqrt{5}}{5}$ , then there exists a noncontextual hidden-variable model for Protocols 1 through 6.

We will prove this by explicitly constructing a noncontextual hidden-variable model that correctly reproduces the outcomes of Protocols 1 through 6 as predicted by quantum mechanics. (An alternative approach would be to apply the General Probabilistic Theory (GPT) formalism [24–26] to KD distributions [32] and then prove the existence of a noncontextual hidden-variable model.) We will use the notation set out in Table I. Any such hidden-variable model must be consistent with the predictions of quantum theory [Eqs. (6)–(8) of the Article]. We thus get the following “Correctness Constraints”:

$$\sum_{\lambda \in \Lambda} \mu_\rho(\lambda) \xi_{\Pi_k}(z|\lambda) = f_k^{(1)}(z), \quad (\text{D1})$$

$$\sum_{\lambda, \lambda' \in \Lambda} \mu_\rho(\lambda) \xi_{P_j}^{\text{WX}}(x, \lambda'|\lambda) \xi_{\Pi_k}(z|\lambda') = f_{j,k}^{(2)}(x, z), \quad (\text{D2})$$

$$\sum_{\lambda, \lambda' \in \Lambda} \mu_\rho(\lambda) \xi_{P_j}^{\text{WY}}(y, \lambda'|\lambda) \xi_{\Pi_k}(z|\lambda') = f_{j,k}^{(3)}(y, z), \quad (\text{D3})$$

where we have chosen the ontic space  $\Lambda$  to be finite, turning the integrals into sums.

To ensure the model is noncontextual, we must ensure the hidden-variable model satisfies the “Noncontextuality Constraints” described in the Article [Eqs. (12)–(15)]. They are

$$\sum_{\lambda' \in \Lambda} \xi_{P_j}^{\text{WX}}(x, \lambda'|\lambda) = (1 - p_m)\frac{1}{2} + p_m \xi_{P_j}(x|\lambda), \quad (\text{D4})$$

$$\sum_{x \in \{\pm 1\}} \xi_{P_j}^{\text{WX}}(x, \lambda'|\lambda) = (1 - p_d)\delta_{\lambda\lambda'} + p_d \Gamma_{D_j}(\lambda'|\lambda), \quad (\text{D5})$$

$$\sum_{\lambda' \in \Lambda} \xi_{P_j}^{\text{WY}}(y, \lambda'|\lambda) = \frac{1}{2}, \quad (\text{D6})$$

$$\sum_{y \in \{\pm 1\}} \xi_{P_j}^{\text{WY}}(y, \lambda'|\lambda) = (1 - p_d)\delta_{\lambda\lambda'} + p_d \Gamma_{D_j}(\lambda'|\lambda). \quad (\text{D7})$$

To construct a noncontextual hidden-variable model, we must choose  $\mu_\rho$ ,  $\xi_{\Pi_k}$ ,  $\xi_{P_j}$ ,  $\xi_{P_j}^{\text{WX}}$ ,  $\xi_{P_j}^{\text{WY}}$ , and  $\Gamma_{D_j}$  such that Eqs. (D1)–(D7) are satisfied for all  $j, k$ . Furthermore, we must ensure that the  $\mu_\rho$ ,  $\xi_{\Pi_k}$ ,  $\xi_{P_j}$ ,  $\xi_{P_j}^{\text{WX}}$ ,  $\xi_{P_j}^{\text{WY}}$ , and  $\Gamma_{D_j}$  we choose are probability distributions in the sense that they are normalized to 1 and take values in  $[0, 1]$ . We will prove that a noncontextual hidden-variable model exists by explicitly providing  $\mu_\rho$ ,  $\xi_{\Pi_k}$ ,  $\xi_{P_j}$ ,  $\xi_{P_j}^{\text{WX}}$ ,  $\xi_{P_j}^{\text{WY}}$ , and  $\Gamma_{D_j}$  such that these constraints are satisfied.

Let us start by choosing our ontic space. We will set  $\Lambda = \mathbb{Z}_d$ . Our intuition will be that the hidden-variable model simply stores what the outcome of the  $B_k$  measurement will be. We thus force  $\xi_{\Pi_k}$  to be outcome deterministic:

$$\xi_{\Pi_k}(z|\lambda) = \begin{cases} \delta_{k\lambda} & \text{if } z = +1 \\ 1 - \delta_{k\lambda} & \text{if } z = -1, \end{cases} \quad (\text{D8})$$

which is clearly a normalized probability distribution. Next, we set  $\mu_\rho$  to be the probability distribution from measuring  $B$ . Thus, using the notation introduced in Appendix A, we set

$$\mu_\rho(\lambda) = p_{+1}^\lambda. \quad (\text{D9})$$

We can now check Eq. (D1) for  $z = +1$  by plugging in Eqs. (D8) and (D9):

$$\sum_{\lambda \in \Lambda} \mu_\rho(\lambda) \xi_{\Pi_k}(+1|\lambda) = \sum_{\lambda \in \mathbb{Z}_d} p_{+1}^\lambda \delta_{k\lambda} \quad (\text{D10})$$

$$= p_{+1}^k \quad (\text{D11})$$

$$= f_k^{(1)}(+1), \quad (\text{D12})$$

which is correct. Furthermore, we may also check Eq. (D1) for  $z = -1$ :

$$\sum_{\lambda \in \Lambda} \mu_\rho(\lambda) \xi_{\Pi_k}(-1|\lambda) = \sum_{\lambda \in \mathbb{Z}_d} p_{+1}^\lambda (1 - \delta_{k\lambda}) \quad (\text{D13})$$

$$= 1 - p_{+1}^k \quad (\text{D14})$$

$$= p_{-1}^k \quad (\text{D15})$$

$$= f_k^{(1)}(-1), \quad (\text{D16})$$

which is also as required. Next, to check Eq. (D2), we must provide  $\xi_{P_j}^{\text{WX}}$ . We use the following ansatz:

$$\xi_{P_j}^{\text{WX}}(x, \lambda'|\lambda) = \delta_{\lambda\lambda'} \left[ \frac{1}{2}(1 - p_d) + \frac{1}{2}x p_m (2\text{Rew}_{j,\lambda'} - 1) \right] + \frac{1}{2}p_d \phi_j(\lambda'|\lambda), \quad (\text{D17})$$

where  $\phi_j$  is a stochastic matrix that we choose such that it is normalized [i.e.,  $\sum_{\lambda'} \phi_j(\lambda'|\lambda) = 1$ ] and that it satisfies

$$\sum_{\lambda} \phi_j(\lambda'|\lambda) p_{+1}^{\lambda} = p_{+1}^{D_{j\lambda'}}. \quad (\text{D18})$$

Such a  $\phi_j$  exists, as we may, for example, set  $\phi_j(\lambda'|\lambda) = p_{+1}^{D_{j\lambda'}}$ . We must now ensure that this ansatz yields a probability distribution. We first check normalization:

$$\begin{aligned} \sum_{\lambda' \in \mathbb{Z}_d, x \in \pm 1} \xi_{P_j}^{\text{WX}}(x, \lambda'|\lambda) &= \sum_{\lambda' \in \mathbb{Z}_d, x \in \pm 1} \left( \delta_{\lambda\lambda'} \left[ \frac{1}{2}(1 - p_d) + \frac{1}{2} x p_m (2\text{Re} w_{j,\lambda'} - 1) \right] + \frac{1}{2} p_d \phi_j(\lambda'|\lambda) \right) \\ &= \sum_{\lambda' \in \mathbb{Z}_d} [\delta_{\lambda\lambda'} (1 - p_d) + p_d \phi_j(\lambda'|\lambda)] \\ &= 1 - p_d + p_d = 1. \end{aligned} \quad (\text{D19})$$

Next, we show that if  $\epsilon < \frac{\sqrt{5}}{5}$ , then  $\xi_{P_j}^{\text{WX}}(x, \lambda'|\lambda) \in [0, 1]$ .

*Lemma D2.* If  $\epsilon < \frac{\sqrt{5}}{5}$ , then  $\xi_{P_j}^{\text{WX}}(x, \lambda'|\lambda)$  as given in Eq. (D17) lies in  $[0, 1]$ .

*Proof.*  $\xi_{P_j}^{\text{WX}}(x, \lambda'|\lambda)$  is clearly real, so it suffices to show that  $\xi_{P_j}^{\text{WX}}(x, \lambda'|\lambda) \geq 0$  since normalization then implies that  $\xi_{P_j}^{\text{WX}}(x, \lambda'|\lambda) \leq 1$ . To show this, note that

$$\begin{aligned} \xi_{P_j}^{\text{WX}}(x, \lambda'|\lambda) &= \delta_{\lambda\lambda'} \left[ \frac{1}{2}(1 - p_d) + \frac{1}{2} x p_m (2\text{Re} w_{j,\lambda'} - 1) \right] + \frac{1}{2} p_d \phi_j(\lambda'|\lambda) \\ &\geq \delta_{\lambda\lambda'} \left[ \frac{1}{2}(1 - p_d) + \frac{1}{2} x p_m (2\text{Re} w_{j,\lambda'} - 1) \right] \\ &\geq \delta_{\lambda\lambda'} \left[ \frac{1}{2} \cos^2 \epsilon - \frac{1}{2} \sin 2\epsilon \right] \\ &= \delta_{\lambda\lambda'} h(\epsilon), \end{aligned} \quad (\text{D20})$$

where we defined  $h(\epsilon) := \frac{1}{2} \cos^2 \epsilon - \frac{1}{2} \sin 2\epsilon$ . By the Mean Value Theorem, there exists a  $c \in [0, \epsilon]$  such that

$$h'(c)\epsilon + h(0) = h(\epsilon). \quad (\text{D21})$$

There exists some phase  $\psi$  such that

$$h'(c) = \frac{1}{2} \sin 2c - \cos 2c = \frac{\sqrt{5}}{2} \sin(2c + \psi) \geq -\frac{\sqrt{5}}{2}. \quad (\text{D22})$$

Substituting Eq. (D21) into Eq. (D20) and using Eq. (D22) and that  $h(0) = \frac{1}{2}$  yields

$$\xi_{P_j}^{\text{WX}}(x, \lambda'|\lambda) = \delta_{\lambda\lambda'} [h'(c)\epsilon + h(0)] \geq \delta_{\lambda\lambda'} \left[ \frac{-\sqrt{5}}{2} \epsilon + \frac{1}{2} \right] \geq 0, \quad (\text{D23})$$

where in the last line we used that  $\epsilon < \frac{\sqrt{5}}{5}$ . This completes the proof.  $\blacksquare$

We have thus found that the ansatz (D17) is a normalized probability distribution. We proceed to check that it indeed satisfies Eq. (D2) for  $z = +1$ :

$$\begin{aligned} \sum_{\lambda, \lambda' \in \Lambda} \mu_{\rho}(\lambda) \xi_{P_j}^{\text{WX}}(x, \lambda'|\lambda) \xi_{\Pi_k}(+1|\lambda') &= \sum_{\lambda, \lambda' \in \mathbb{Z}_d} p_{+1}^{\lambda} \xi_{P_j}^{\text{WX}}(x, \lambda'|\lambda) \delta_{k\lambda'} \\ &= \sum_{\lambda \in \mathbb{Z}_d} p_{+1}^{\lambda} \xi_{P_j}^{\text{WX}}(x, k|\lambda) \\ &= \sum_{\lambda \in \mathbb{Z}_d} p_{+1}^{\lambda} \left\{ \delta_{\lambda k} \left[ \frac{1}{2}(1 - p_d) + \frac{1}{2} x p_m (2\text{Re} w_{j,k} - 1) \right] + \frac{1}{2} p_d \phi_j(k|\lambda) \right\} \\ &= \frac{1}{2} (1 - p_d) p_{+1}^k + \frac{1}{2} x p_m p_{+1}^k (2\text{Re} w_{j,k} - 1) + \frac{1}{2} p_d \sum_{\lambda \in \mathbb{Z}_d} \phi_j(k|\lambda) p_{+1}^{\lambda} \\ &= \frac{1}{2} (1 - p_d) p_{+1}^k + \frac{1}{2} x p_m p_{+1}^k (2\text{Re} w_{j,k} - 1) + \frac{1}{2} p_d p_{+1}^{D_{j,k}} \\ &= f_{j,k}^{(2)}(x, +1), \end{aligned} \quad (\text{D24})$$

which is correct. Similarly, we may check Eq. (D2) for  $z = -1$ :

$$\begin{aligned}
\sum_{\lambda, \lambda' \in \Lambda} \mu_\rho(\lambda) \xi_{P_j}^{\text{WX}}(x, \lambda' | \lambda) \xi_{\Pi_k}(-1 | \lambda') &= \sum_{\lambda, \lambda' \in \mathbb{Z}_d} p_S^\lambda \xi_{P_j}^{\text{WX}}(x, \lambda' | \lambda) (1 - \delta_{k\lambda'}) \\
&= \sum_{\lambda \in \mathbb{Z}_d, \lambda' \in \mathbb{Z}_d \setminus \{k\}} p_S^\lambda \xi_{P_j}^{\text{WX}}(x, \lambda' | \lambda) \\
&= \sum_{\lambda' \in \mathbb{Z}_d \setminus \{k\}} \frac{1}{2} (1 - p_d) p_S^{\lambda'} + \frac{1}{2} x p_m p_S^{\lambda'} (2 \text{Rew}_{j\lambda'} - 1) + \frac{1}{2} p_d p_S^{D_{j\lambda'}} \\
&= \frac{1}{2} (1 - p_d) p_{-1}^k + \frac{1}{2} x p_m p_{-1}^k \left( 2 \frac{\sum_{\lambda' \in \mathbb{Z}_d \setminus \{k\}} p_S^{\lambda'} \text{Rew}_{j\lambda'}}{p_{-1}^k} - 1 \right) + \frac{1}{2} p_d p_{-1}^{D_{jk}} \\
&= \frac{1}{2} (1 - p_d) p_{-1}^k + \frac{1}{2} x p_m p_{-1}^k (2g_{-1}(\text{Rew}_{jk}) - 1) + \frac{1}{2} p_d p_{-1}^{D_{jk}} \\
&= f_{j,k}^{(2)}(x, -1),
\end{aligned} \tag{D25}$$

which is also correct. We thus conclude that our ansatz satisfies Eq. (D2).

We move on to checking the Noncontextuality Constraints involving  $\xi_{P_j}^{\text{WX}}$ . We start with Eq. (D4) for  $x = +1$ :

$$\begin{aligned}
\sum_{\lambda' \in \Lambda} \xi_{P_j}^{\text{WX}}(+1, \lambda' | \lambda) &= \sum_{\lambda' \in \mathbb{Z}_d} \left( \delta_{\lambda\lambda'} \left[ \frac{1}{2} (1 - p_d) + \frac{1}{2} p_m (2 \text{Rew}_{j\lambda'} - 1) \right] + \frac{1}{2} p_d \phi_j(\lambda' | \lambda) \right) \\
&= \frac{1}{2} (1 - p_d) + \frac{1}{2} p_m (2 \text{Rew}_{j,\lambda} - 1) + \frac{1}{2} p_d \\
&= (1 - p_m) \frac{1}{2} + p_m \text{Rew}_{j,\lambda}.
\end{aligned} \tag{D26}$$

Comparing this with the right-hand side of Eq. (D4) for  $x = +1$ , we see that we must set  $\xi_{P_j}(+1 | \lambda) = \text{Rew}_{j,\lambda}$ . This yields probabilities in  $[0, 1]$  since the fact that  $\rho$  is KD positive implies that the weak values lie in  $[0, 1]$  by Eq. (A7). Furthermore, we fix the  $\xi_{P_j}(+1 | \lambda)$  via normalization:  $\xi_{P_j}(-1 | \lambda) = 1 - \text{Rew}_{j,\lambda}$ . We can now also check Eq. (D4) for  $x = -1$ :

$$\begin{aligned}
\sum_{\lambda' \in \Lambda} \xi_{P_j}^{\text{WX}}(-1, \lambda' | \lambda) &= \sum_{\lambda' \in \mathbb{Z}_d} \left( \delta_{\lambda\lambda'} \left[ \frac{1}{2} (1 - p_d) - \frac{1}{2} p_m (2 \text{Rew}_{j\lambda'} - 1) \right] + \frac{1}{2} p_d \phi_j(\lambda' | \lambda) \right) \\
&= \frac{1}{2} (1 - p_d) - \frac{1}{2} p_m (2 \text{Rew}_{j,\lambda} - 1) + \frac{1}{2} p_d \\
&= (1 - p_m) \frac{1}{2} - p_m (1 - \text{Rew}_{j,\lambda}) \\
&= (1 - p_m) \frac{1}{2} - p_m \xi_{P_j}(-1 | \lambda),
\end{aligned} \tag{D27}$$

as required. We thus conclude that Noncontextuality Constraint (D4) is satisfied. We move on to show that Noncontextuality Constraint (D5) is satisfied. We calculate its left-hand side for this ansatz:

$$\begin{aligned}
\sum_{x \in \{\pm 1\}} \xi_{P_j}^{\text{WX}}(x, \lambda' | \lambda) &= \sum_{x \in \{\pm 1\}} \left( \delta_{\lambda\lambda'} \left[ \frac{1}{2} (1 - p_d) + \frac{1}{2} x p_m (2 \text{Rew}_{j\lambda'} - 1) \right] + \frac{1}{2} p_d \phi_j(\lambda' | \lambda) \right) \\
&= (1 - p_d) \delta_{\lambda\lambda'} + p_d \phi_j(\lambda' | \lambda).
\end{aligned} \tag{D28}$$

Comparing this result to the right-hand side of Eq. (D5), we find that we must set  $\Gamma_{D_j}(\lambda' | \lambda) = \phi_j(\lambda' | \lambda)$ . By construction,  $\phi_j(\lambda' | \lambda)$  is a normalized stochastic matrix with all entries in  $[0, 1]$ . Hence, this assigns  $\Gamma_{D_j}(\lambda' | \lambda)$  to a normalized probability distribution between 0 and 1. Hence, Eq. (D5) is satisfied as well.

We have thus demonstrated that we have adequate solutions to the equations involving  $\xi_{P_j}^{\text{WX}}$ . We solve the equations involving  $\xi_{P_j}^{\text{WY}}$  similarly by providing an ansatz,

$$\xi_{P_j}^{\text{WY}}(y, \lambda' | \lambda) = \delta_{\lambda\lambda'} \left[ \frac{1}{2} (1 - p_d) + y p_m \text{Imw}_{j,\lambda'} \right] + \frac{1}{2} p_d \phi_j(\lambda' | \lambda), \tag{D29}$$

where  $\phi_j$  is chosen as before. Showing that this ansatz is normalized and only takes values in  $[0,1]$  is done in exactly the same way as for  $\xi_{P_j}^{\text{WX}}$ , and we omit it here. First, we check Eq. (D3) for  $z = +1$ :

$$\begin{aligned}
\sum_{\lambda, \lambda' \in \Lambda} \mu_\rho(\lambda) \xi_{P_j}^{\text{WY}}(y, \lambda'|\lambda) \xi_{\Pi_k}(-1|\lambda') &= \sum_{\lambda, \lambda' \in \mathbb{Z}_d} p_{+1}^\lambda \xi_{P_j}^{\text{WY}}(y, \lambda'|\lambda) \delta_{\lambda'k} \\
&= \sum_{\lambda \in \mathbb{Z}_d} p_{+1}^\lambda \xi_{P_j}^{\text{WY}}(y, k|\lambda) \\
&= \sum_{\lambda \in \mathbb{Z}_d} \left( p_{+1}^\lambda \delta_{\lambda k} \left[ \frac{1}{2}(1 - p_d) + y p_m \text{Im} w_{j,k} \right] + \frac{1}{2} p_d \phi_j(k|\lambda) \right) \\
&= \frac{1}{2}(1 - p_d) p_{+1}^k + y p_m p_{+1}^k \text{Im} w_{j,k} + \frac{1}{2} p_d \sum_{\lambda \in \mathbb{Z}_d} p_{+1}^\lambda \phi_j(k|\lambda) \\
&= \frac{1}{2}(1 - p_d) p_{+1}^k + y p_m p_{+1}^k \text{Im} w_{j,k} + \frac{1}{2} p_d p_{+1}^{D_{j,k}} \\
&= f_{j,k}^{(3)}(y, +1).
\end{aligned} \tag{D30}$$

Hence, Eq. (D3) is satisfied for  $z = +1$ . We also check Eq. (D6) for  $z = -1$ :

$$\begin{aligned}
\sum_{\lambda, \lambda' \in \Lambda} \mu_\rho(\lambda) \xi_{P_j}^{\text{WY}}(y, \lambda'|\lambda) \xi_{\Pi_k}(-1|\lambda') &= \sum_{\lambda, \lambda' \in \mathbb{Z}_d} p_{+1}^\lambda \xi_{P_j}^{\text{WY}}(y, \lambda'|\lambda) (1 - \delta_{\lambda'k}) \\
&= \sum_{\lambda \in \mathbb{Z}_d, \lambda' \in \mathbb{Z}_d \setminus \{k\}} p_{+1}^\lambda \xi_{P_j}^{\text{WY}}(y, \lambda'|\lambda) \\
&= \sum_{\lambda' \in \mathbb{Z}_d \setminus \{k\}} \left( \frac{1}{2}(1 - p_d) p_{+1}^{\lambda'} + y p_m p_{+1}^{\lambda'} \text{Im} w_{j,\lambda'} + \frac{1}{2} p_d p_{+1}^{D_{j,\lambda'}} \right) \\
&= \frac{1}{2}(1 - p_d) p_{-1}^k + y p_m p_{-1}^k \frac{\sum_{\lambda' \in \mathbb{Z}_d \setminus \{k\}} p_{+1}^{\lambda'} \text{Im} w_{j,\lambda'}}{p_{-1}^k} + \frac{1}{2} p_d p_{-1}^{D_{j,k}} \\
&= \frac{1}{2}(1 - p_d) p_{-1}^k + y p_m p_{-1}^k g_{-1}(\text{Im} w_{j,\lambda'}) + \frac{1}{2} p_d p_{-1}^{D_{j,k}} \\
&= f_{j,k}^{(3)}(y, -1).
\end{aligned} \tag{D31}$$

Hence, Eq. (D6) is also satisfied for  $z = -1$ . Last, we will need to check the two Nontextuality Constraints associated with  $\xi_{P_j}^{\text{WY}}$ . We start with Eq. (D6):

$$\begin{aligned}
\sum_{\lambda' \in \Lambda} \xi_{P_j}^{\text{WY}}(y, \lambda'|\lambda) &= \sum_{\lambda' \in \mathbb{Z}_d} \left( \delta_{\lambda\lambda'} \left[ \frac{1}{2}(1 - p_d) + y p_m \text{Im} w_{j,\lambda'} \right] + \frac{1}{2} p_d \phi_j(\lambda'|\lambda) \right) \\
&= \sum_{\lambda' \in \mathbb{Z}_d} \delta_{\lambda\lambda'} \frac{1}{2}(1 - p_d) + \frac{1}{2} p_d \phi_j(\lambda'|\lambda) \\
&= \frac{1}{2}(1 - p_d) + \frac{1}{2} p_d = \frac{1}{2},
\end{aligned} \tag{D32}$$

where to go from the first to the second line we used that by Eq. (A7) the fact that  $\rho$  is KD positive implies that  $\text{Im} w_{j,\lambda'} = 0$ . Hence, Eq. (D6) is satisfied. Last, we need to check Eq. (D7):

$$\begin{aligned}
\sum_{y \in \{\pm 1\}} \xi_{P_j}^{\text{WY}}(y, \lambda'|\lambda) &= \sum_{y \in \{\pm 1\}} \left( \delta_{\lambda\lambda'} \left[ \frac{1}{2}(1 - p_d) + y p_m \text{Im} w_{j,\lambda'} \right] + \frac{1}{2} p_d \phi_j(\lambda'|\lambda) \right) \\
&= (1 - p_d) \delta_{\lambda\lambda'} + p_d \phi_j(\lambda'|\lambda) \\
&= (1 - p_d) \delta_{\lambda\lambda'} + p_d \Gamma_{D_j}(\lambda'|\lambda),
\end{aligned} \tag{D33}$$

where we use that we chose  $\phi_j$  earlier to be  $\Gamma_{D_j}$ . Hence, we have explicitly demonstrated that if the state is KD positive and  $\epsilon$  is sufficiently small, that then there exists a noncontextual hidden-variable model for Protocols 1 through 6.

### APPENDIX E: EXPLICIT EXAMPLE FOR A SPIN-1 SYSTEM

In this Appendix, we construct an explicit experiment on a spin-1 system of Hilbert space dimension 3, the lowest dimension in which exotic states exist [37]. The contextuality of this experiment may be verified by noncontextual data. Following the example given in Ref. [37], we consider a spin-1 system. We choose the observable  $A$  to be the  $z$  component of the spin of the particle, and we choose observable  $B$  to be the component of the spin of the particle along the axis  $\hat{n} = -(2, 2, 1)^T/3$ . We will denote the spin  $j \in \{-1, 0, 1\}$  eigenstate of the  $z$  component of spin by the vector  $|j\rangle$ .

We first consider the experiment from Alice's perspective. Reference [37] showed that the state given in the eigenbasis of  $A$  by the density matrix

$$\rho_\star = \frac{1}{15} \begin{pmatrix} 5 & 1 & 1 \\ 1 & 5 & 1 \\ 1 & 1 & 5 \end{pmatrix} \quad (\text{E1})$$

is an exotic state for observables  $A$  and  $B$ . Alice may prepare this state by, for example, using the sequence of  $N = 15$  quantum states containing  $|-1\rangle$  four times,  $|0\rangle$  four times,  $|1\rangle$  four times, and  $|+\rangle$  three times, where only  $|+\rangle := (|0\rangle + |1\rangle + |2\rangle)/\sqrt{3}$  is KD nonpositive. One may easily verify that the mixing of this sequence indeed prepares  $\rho_\star$ :

$$\rho_\star = \frac{3}{15}|+\rangle\langle+| + \frac{4}{15}|-1\rangle\langle-1| + \frac{4}{15}|0\rangle\langle 0| + \frac{4}{15}|1\rangle\langle 1|. \quad (\text{E2})$$

Concretely, Alice prepares this sequence of 15 quantum states and chooses a secret permutation of this sequence. She then sends the states to Bob one by one, in the order of this permutation. She repeats this process many times, rechoosing the permutation each time.

We now consider the experiment from Bob's perspective. We start by describing how Bob can implement all the procedures listed in Table I. First, Bob needs to implement measurements of the observables  $P_j$  and  $\Pi_k$ . Since the observable  $A$  is the component of the spin of the particle along the  $z$  axis, the measurement of  $P_j$  corresponds to checking if the particle has spin  $j$  along the  $z$  axis. Similarly, since the observable  $B$  is the component of spin along the  $\hat{n}$  axis, the measurement of  $\Pi_k$  corresponds to checking if the particle has spin  $k$  along the  $\hat{n}$  axis.

Bob also has to implement the quantum channel  $\mathcal{M}^{D_j}(\rho) := D_j \rho D_j^\dagger$ , where  $D_j := 2P_j - I$ . A straightforward calculation shows that  $D_j$  is equivalent to a unitary that acts as identity on every basis state of the eigenbasis of  $A$ , except for the state  $|j\rangle$  on which it induces a phase of  $-1$ . Bob can therefore implement  $\mathcal{M}^{D_j}$  by applying this unitary.

Last, Bob has to implement the weak measurements  $W_{P_j}^{(X)}$  and  $W_{P_j}^{(Y)}$ . Bob can do this by implementing the quantum circuit given in Fig. 3. Note that Bob needs to pick some  $\epsilon$  as the strength parameter of the weak measurement. He announces his choice on the public channel.

Using these explicit procedures, Bob can implement all six protocols. For every state Bob receives from Alice, he picks a random protocol and uses it on the state. After each measurement, he announces what protocol he used and his result on the public channel.

After many repetitions, Bob will be able to reconstruct the probability distributions  $f_k^{(1)}(z)$ ,  $f_{j,k}^{(1)}(x, z)$ ,  $f_{j,k}^{(3)}(y, z)$ ,  $f_j^{(4)}(x)$ ,  $f_j^{(5)}(y)$ , and  $f_{j,k}^{(6)}(z)$ . Using Eqs. (7) and (8) of the Article, Bob can reconstruct the real and imaginary parts of the KD distribution. Using this KD distribution, Bob can then calculate the KD nonpositivity  $\mathcal{N}(\rho_\star)$ . Since  $\rho_\star$  was chosen by Alice to be KD positive, Bob will find  $\mathcal{N}(\rho_\star) = 0$  to within some experimental error. This will allow Bob to conclude that there exists a noncontextual model that is consistent with his data.

Using the KD distribution he reconstructed, Bob may also reconstruct the quantum state  $\rho_\star$ , allowing him to conclude that the state he was given was indeed exotic. Using the mathematical framework in Appendix F, Bob can use  $\rho_\star$  to calculate a  $\delta$  such that every pure-state decomposition of  $\rho_\star$  has at least one state  $\psi_-$  such that  $\mathcal{N}(\psi_-) > \delta$ . Using the same logic as in the Article, if Bob finds that  $\delta > 3d^2\epsilon$ , then he has verified that Alice has witnessed contextuality.

### APPENDIX F: BOUNDING THE NEGATIVITY OF DECOMPOSITIONS OF EXOTIC STATES

The main result of the Article depends on the claim that every decomposition of an exotic state into pure states has at least one pure state with negativity bounded away from zero. We prove this claim in this Appendix. More precisely, we prove the following statement:

*Lemma F1 (Every decomposition of an exotic state has a state with bounded negativity).* Let  $\rho_\star \in \mathcal{E}_{\text{KD}+}^{\text{exot}}$  be an exotic state. There exists a  $\delta > 0$  such that for every decomposition of  $\rho_\star$  into pure states,  $\rho_\star = \sum_i p_i \psi_i$ , there exists at least one pure state  $\psi_- \in (\psi_i)_i$  such that  $\mathcal{N}(\psi_-) > \delta$ .

To prove this, we aim to construct a set containing all the pure states with negativity bounded away from zero. We do this as follows. We have that  $\text{conv}(\mathcal{E}_{\text{KD}+}^{\text{pure}})$  and  $\{\rho_\star\}$  are both closed and compact sets. The hyperplane separation theorem implies that there exists a Hermitian operator  $H$  and real constants  $d > c$  such that (i)  $\text{Tr}(H\rho_\star) = d$  and (ii)  $\forall \rho \in \text{conv}(\mathcal{E}_{\text{KD}+}^{\text{pure}})$ ,  $\text{Tr}(H\rho) \leq c$ . We now introduce the set

$$\mathcal{A} := \{\rho \in \mathcal{D}(\mathbb{C}^d) \mid \rho \text{ is pure and } \text{Tr}(H\rho) \geq d\},$$

where  $\mathcal{D}(\mathbb{C}^d)$  denotes the set of density operators on  $\mathbb{C}^d$ , and we prove that this set has the following desirable properties.

*Lemma F2 (Properties of  $\mathcal{A}$ ).* The set  $\mathcal{A}$  has the following properties:

(i) For every decomposition of an exotic state  $\rho_\star$  into pure states,  $\rho_\star = \sum_i p_i \psi_i$ , there exists at least one pure state  $\psi_i$  such that  $\psi_i \in \mathcal{A}$ .

(ii) There exists a real constant  $\delta > 0$  such that for all  $\rho \in \mathcal{A}$ ,  $\mathcal{N}(\rho) \geq \delta$ .

*Proof.* Property (i) by contradiction. Suppose that for all  $i$ ,  $\psi_i \notin \mathcal{A}$ . We then calculate

$$\text{Tr}(H\rho_\star) = \sum_i p_i \text{Tr}(H\psi_i) < \sum_i p_i d = d, \quad (\text{F1})$$

where we used the defining property of  $\mathcal{A}$  in the inequality. However, from the definition of  $H$ , we know that  $\text{Tr}(H\rho_\star) = d$ . We thus find a contradiction, establishing the result.

Property (ii):  $\mathcal{A}$  is closed and compact, and  $\mathcal{N}$  is continuous. The Extreme Value Theorem then implies that  $\mathcal{N}$

is bounded on  $\mathcal{A}$ , and that  $\mathcal{N}$  attains this bound at some  $\rho_{\min} \in \mathcal{A}$ .

Now suppose that  $\mathcal{N}(\rho_{\min}) = 0$ . This implies that  $\rho_{\min}$  is KD positive. All states in  $\mathcal{A}$  are pure, so we have that  $\rho_{\min} \in \mathcal{E}_{\text{KD}^+}^{\text{pure}}$ . From the definition of  $H$ , this yields

$$\text{Tr}(H\rho_{\min}) \leq c < d. \quad (\text{F2})$$

However, since  $\rho_{\min} \in \mathcal{A}$ , we have that  $\text{Tr}(H\rho_{\min}) \geq d$  as well. We thus find a contradiction, forcing us to conclude that  $\mathcal{N}(\rho_{\min}) \neq 0$ . We may thus set  $\delta = \mathcal{N}(\rho_{\min}) > 0$  to obtain the desired result. ■

The proof of Lemma F 1 then straightforwardly follows from Lemma F 2.

- 
- [1] S. Aaronson and D. Gottesman, Improved simulation of stabilizer circuits, *Phys. Rev. A* **70**, 052328 (2004).
  - [2] C. Weedbrook, S. Pirandola, R. García-Patrón, N. J. Cerf, T. C. Ralph, J. H. Shapiro, and S. Lloyd, Gaussian quantum information, *Rev. Mod. Phys.* **84**, 621 (2012).
  - [3] V. Veitch, C. Ferrie, D. Gross, and J. Emerson, Negative quasiprobability as a resource for quantum computation, *New J. Phys.* **14**, 113011 (2012).
  - [4] V. Veitch, S. A. H. Mousavian, D. Gottesman, and J. Emerson, The resource theory of stabilizer quantum computation, *New J. Phys.* **16**, 013009 (2014).
  - [5] R. Raussendorf, J. Bermejo-Vega, E. Tyhurst, C. Okay, and M. Zurek, Phase-space-simulation method for quantum computation with magic states on qubits, *Phys. Rev. A* **101**, 012350 (2020).
  - [6] V. Giovannetti, S. Lloyd, and L. Maccone, Quantum metrology, *Phys. Rev. Lett.* **96**, 010401 (2006).
  - [7] V. Giovannetti, S. Lloyd, and L. Maccone, Advances in quantum metrology, *Nat. Photon.* **5**, 222 (2011).
  - [8] L. Maccone, Intuitive reason for the usefulness of entanglement in quantum metrology, *Phys. Rev. A* **88**, 042109 (2013).
  - [9] D. R. M. Arvidsson-Shukur, N. Yünger Halpern, H. V. Lepage, A. A. Lasek, C. H. W. Barnes, and S. Lloyd, Quantum advantage in postselected metrology, *Nat. Commun.* **11**, 3775 (2020).
  - [10] N. Lupu-Gladstein, Y. B. Yilmaz, D. R. M. Arvidsson-Shukur, A. Brodutch, A. O. Pang, A. M. Steinberg, and N. Y. Halpern, Negative quasiprobabilities enhance phase estimation in quantum-optics experiment, *Phys. Rev. Lett.* **128**, 220504 (2022).
  - [11] E. S. Simon Kochen, The problem of hidden variables in quantum mechanics, *Indiana Univ. Math. J.* **17**, 59 (1967).
  - [12] E. Specker, Die logik nicht gleichzeitig entscheidbarer Aussagen, *Dialectica* **14**, 239 (1960).
  - [13] C. Budroni, A. Cabello, O. Gühne, M. Kleinmann, and J.-A. Larsson, Kochen-specker contextuality, *Rev. Mod. Phys.* **94**, 045007 (2022).
  - [14] S. Abramsky and A. Brandenburger, The sheaf-theoretic structure of non-locality and contextuality, *New J. Phys.* **13**, 113036 (2011).
  - [15] E. N. Dzhafarov, V. H. Cervantes, and J. V. Kujala, Contextuality in canonical systems of random variables, *Philos. Trans. R. Soc. A* **375**, 20160389 (2017).
  - [16] E. N. Dzhafarov, On joint distributions, counterfactual values and hidden variables in understanding contextuality, *Philos. Trans. R. Soc. A* **377**, 20190144 (2019).
  - [17] J. V. Kujala and E. N. Dzhafarov, Measures of contextuality and non-contextuality, *Philos. Trans. R. Soc. A* **377**, 20190149 (2019).
  - [18] R. W. Spekkens, Contextuality for preparations, transformations, and unsharp measurements, *Phys. Rev. A* **71**, 052108 (2005).
  - [19] R. W. Spekkens, Negativity and contextuality are equivalent notions of nonclassicality, *Phys. Rev. Lett.* **101**, 020401 (2008).
  - [20] M. Leifer, Is the quantum state real? An extended review of psi-ontology theorems, *Quanta* **3**, 67 (2014).
  - [21] M. D. Mazurek, M. F. Pusey, R. Kunjwal, K. J. Resch, and R. W. Spekkens, An experimental test of noncontextuality without unphysical idealizations, *Nat. Commun.* **7**, ncomms11780 (2016).
  - [22] D. Schmid and R. W. Spekkens, Contextual advantage for state discrimination, *Phys. Rev. X* **8**, 011015 (2018).
  - [23] M. Lostaglio, Certifying quantum signatures in thermodynamics and metrology via contextuality of quantum linear response, *Phys. Rev. Lett.* **125**, 230603 (2020).
  - [24] D. Schmid, J. H. Selby, E. Wolfe, R. Kunjwal, and R. W. Spekkens, Characterization of noncontextuality in the framework of generalized probabilistic theories, *PRX Quantum* **2**, 010331 (2021).
  - [25] J. H. Selby, D. Schmid, E. Wolfe, A. B. Sainz, R. Kunjwal, and R. W. Spekkens, Accessible fragments of generalized probabilistic theories, cone equivalence, and applications to witnessing nonclassicality, *Phys. Rev. A* **107**, 062203 (2023).
  - [26] D. Schmid, J. H. Selby, M. F. Pusey, and R. W. Spekkens, A structure theorem for generalized-noncontextual ontological models, *Quantum* **8**, 1283 (2024).
  - [27] J. S. Bell, On the Einstein Podolsky Rosen paradox, *Phys. Phys. Fiz.* **1**, 195 (1964).
  - [28] J. G. Kirkwood, Quantum statistics of almost classical assemblies, *Phys. Rev.* **44**, 31 (1933).
  - [29] P. A. M. Dirac, On the analogy between classical and quantum mechanics, *Rev. Mod. Phys.* **17**, 195 (1945).
  - [30] D. R. M. Arvidsson-Shukur, W. F. Braasch, Jr., S. De Bievre, J. Dressel, A. N. Jordan, C. Langrenze, M. Lostaglio, J. S. Lundeen, and N. Yünger Halpern, Properties and applications of the Kirkwood-Dirac distribution *New J. Phys.* **26**, 121201 (2024).
  - [31] M. Lostaglio, A. Belenchia, A. Levy, S. Hernández-Gómez, N. Fabbri, and S. Gherardini, Kirkwood-Dirac quasiprobability approach to the statistics of incompatible observables, *Quantum* **7**, 1128 (2023).
  - [32] D. Schmid, R. D. Baldijão, Y. Ying, R. Wagner, and J. H. Selby, Kirkwood-Dirac representations beyond quantum states and their relation to noncontextuality, *Phys. Rev. A* **110**, 052206 (2024).
  - [33] J. R. González Alonso, N. Yünger Halpern, and J. Dressel, Out-of-time-ordered-correlator quasiprobabilities robustly witness scrambling, *Phys. Rev. Lett.* **122**, 040404 (2019).

- [34] D. Arvidsson-Shukur, J. Chevalier Drori, and N. Yunger Halpern, Conditions tighter than noncommutation needed for nonclassicality, *J. Phys. A: Math. Theor.* **54**, 284001 (2021).
- [35] S. De Bièvre, Complete incompatibility, support uncertainty, and Kirkwood-Dirac nonclassicality, *Phys. Rev. Lett.* **127**, 190404 (2021).
- [36] C. Langrenez, W. Salmon, S. D. Bièvre, J. J. Thio, C. K. Long, and D. R. M. Arvidsson-Shukur, The set of Kirkwood-Dirac positive states is almost always minimal, [arXiv:2405.17557](https://arxiv.org/abs/2405.17557).
- [37] C. Langrenez, D. R. M. Arvidsson-Shukur, and S. De Bièvre, Characterizing the geometry of the Kirkwood-Dirac-positive states, *J. Math. Phys.* **65**, 072201 (2024).
- [38] Y. Aharonov, D. Z. Albert, and L. Vaidman, How the result of a measurement of a component of the spin of a spin-1/2 particle can turn out to be 100, *Phys. Rev. Lett.* **60**, 1351 (1988).
- [39] I. M. Duck, P. M. Stevenson, and E. C. G. Sudarshan, The sense in which a “weak measurement” of a spin-1/2 particle’s spin component yields a value 100, *Phys. Rev. D* **40**, 2112 (1989).
- [40] R. Jozsa, Complex weak values in quantum measurement, *Phys. Rev. A* **76**, 044103 (2007).
- [41] J. Dressel, M. Malik, F. M. Miatto, A. N. Jordan, and R. W. Boyd, Colloquium: Understanding quantum weak values: Basics and applications, *Rev. Mod. Phys.* **86**, 307 (2014).
- [42] R. Wagner, Z. Schwartzman-Nowik, I. L. Paiva, A. Te’eni, A. Ruiz-Molero, R. S. Barbosa, E. Cohen, and E. F. Galvão, Quantum circuits for measuring weak values, Kirkwood-Dirac quasiprobability distributions, and state spectra, *Quantum Sci. Technol.* **9**, 015030 (2024).
- [43] M. F. Pusey, Anomalous weak values are proofs of contextuality, *Phys. Rev. Lett.* **113**, 200401 (2014).
- [44] R. Kunjwal, M. Lostaglio, and M. F. Pusey, Anomalous weak values and contextuality: Robustness, tightness, and imaginary parts, *Phys. Rev. A* **100**, 042116 (2019).
- [45] R. Wagner and E. F. Galvão, Simple proof that anomalous weak values require coherence, *Phys. Rev. A* **108**, L040202 (2023).
- [46] V. Vedral, *Introduction to Quantum Information Science* (Oxford University Press, Oxford, 2006).
- [47] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Quantum entanglement, *Rev. Mod. Phys.* **81**, 865 (2009).
- [48] R. F. Werner, Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model, *Phys. Rev. A* **40**, 4277 (1989).
- [49] H. M. Wiseman, S. J. Jones, and A. C. Doherty, Steering, entanglement, nonlocality, and the Einstein-Podolsky-Rosen paradox, *Phys. Rev. Lett.* **98**, 140402 (2007).
- [50] S. A. Hill and W. K. Wootters, Entanglement of a pair of quantum bits, *Phys. Rev. Lett.* **78**, 5022 (1997).
- [51] W. K. Wootters, Entanglement of formation of an arbitrary state of two qubits, *Phys. Rev. Lett.* **80**, 2245 (1998).
- [52] C. Langrenez, S. D. Bièvre, and D. R. M. Arvidsson-Shukur, Convex roofs witnessing Kirkwood-Dirac nonpositivity *J. Phys. A: Math. Theor.* **58**, 215301 (2025).