

Contextuality Can be Verified with Noncontextual Experiments

Jonathan J. Thio,¹ Wilfred Salmon,^{2,3} Crispin H. W. Barnes,¹
Stephan De Bièvre,⁴ and David R. M. Arvidsson-Shukur³

¹*Cavendish Lab., Department of Physics, Univ. of Cambridge, Cambridge CB3 0HE, United Kingdom*

²*DAMTP, Centre for Mathematical Sciences, Univ. of Cambridge, CB30WA, United Kingdom*

³*Hitachi Cambridge Laboratory, J. J. Thomson Avenue, Cambridge CB3 0HE, United Kingdom*

⁴*Univ. Lille, CNRS, Inria, UMR 8524, Laboratoire Paul Painlevé, F-59000 Lille, France*

(Dated: December 3, 2024)

We uncover new 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 disc. 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.

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 [11–16]. 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 Λ of hidden variables. 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 [11]. However, the hidden-variable model may have nonclassical features, for example, nonlocality [17]. We capture nonclassicality using the notion of generalized contextuality [11]. 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 con-

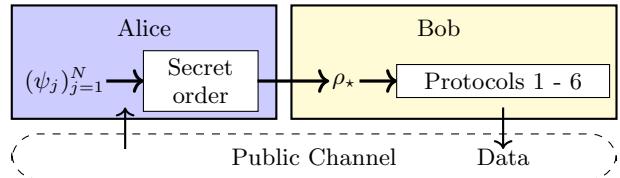


FIG. 1. Noncontextual Experiment that Signals Contextuality. Alice sends the sequence $(\psi_j)_{j=1}^N$ of pure states to Bob without disclosing the order. Bob effectively receives a mixed state $\rho_\star = \frac{1}{N} \sum_{j=1}^N \psi_j$. Bob performs Procedures 1 to 6 (outlined in the main text).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

struct an experiment [Fig. 1], the contextuality of which can be verified with noncontextual procedures. This experiment relies on measurements of a Kirkwood-Dirac (KD) distribution.

KD Distributions:— The KD distributions [18–21] 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 ρ 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. P_j and Π_k are the rank-1 projectors onto the corresponding eigenspaces. The KD distribution of a state ρ is

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

Throughout this work, we assume that $P_j \neq \Pi_k$ for all j and k . Then, Q is invertible: Knowledge of $Q(\rho)$ enables an informationally complete reconstruction of ρ [20].

The KD distribution is not a proper probability distribution as its entries may lie in the complex unit disc. 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 ρ is KD-positive iff $\mathcal{N}(\rho) = 0$ [20, 22]. For example, if ρ equals an eigenstate of A or B , then $\mathcal{N}(\rho) = 0$. However, for pure states, if ρ has nonzero overlap with sufficiently many of the eigenstates of A and B , then ρ is KD-nonpositive: $\mathcal{N}(\rho) > 0$ [23, 24]. It was recently shown, that for almost all choices of A and B , all KD-positive states are convex mixtures of A and B 's basis states [25]. 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 [26]. We call such states ‘exotic’, and we denote the set containing them by $\mathcal{E}_{KD+}^{\text{exot}}$. Exotic states will play the central role in what follows.

Measurements of $Q(\rho)$:—To measure $Q(\rho)$, one can perform a series of protocols on ρ , involving the projective measurement of P_j and Π_k , and the so-called ‘weak measurement’ [27–30] of P_j . The weak measurement involves coupling weakly the A observable of ρ to a qubit ancilla, which later is measured in the X or Y eigenbasis. The strength of the coupling is given by the parameter $\epsilon \ll 1$. The exact implementation of the weak measurement is given in Note I of the Supplementary Material. For our purposes, it suffices to note that the X - and Y -type weak measurements of P_j are represented by the Kraus operators

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

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

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 ρ . Protocols 1 to 3 measure the ‘weak values’ [27–30] of P_j . Protocols 4 through 6 are designed to establish the connection to contextuality, outlined below.

1. Measure Π_k on the system (returning outcome z).
2. First, perform an X -type weak measurement of P_j (returning outcome x) and then measure Π_k (returning outcome z).

3. First perform an Y -type weak measurement of P_j (returning outcome y) and then measure Π_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 Π_k (returning outcome z).

These six protocols include repetition over all j and k . We denote the outcome-probability distributions for these protocols by $f_k^{(1)}(z)$, $f_{j,k}^{(1)}(x, z)$, $f_{j,k}^{(3)}(y, z)$, $f_j^{(4)}(x)$, $f^{(5)}(y)$ 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

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

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

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

Quantum theory allows us to express the outcome-probability distributions of Protocols (9) - (11) in terms of the outcome-probability distributions of protocols (6) - (8):

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

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

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

(Detailed calculations are given in Note I of the Supplementary Material.) 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), (7) and (8).

KD-nonpositivity as a faithful witness of 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:

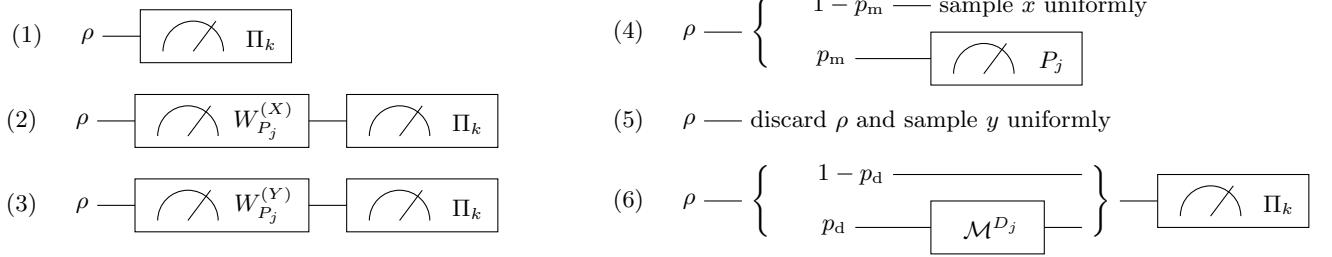


FIG. 2. Experimental Protocols. Upon receiving a state ρ from Alice, Bob randomly implements one of six protocols [see Fig. 1]. The boxes with P_j and Π_k represent projective measurements. The boxes labeled by $W_{P_j}^{(X)}$ and $W_{P_j}^{(Y)}$ represent X -type and Y -type weak measurements of P_j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

(i) a Y -type weak measurement of P_j followed by a discarding of the system, and (ii) entirely ignoring the system and picking an outcome uniformly at random. Thus, in any noncontextual hidden-variable model, these two procedures are represented by the same probability distribution:

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

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

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

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

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

Here, $\xi_{P_j}^{WX}(x, \lambda'|\lambda)$ and $\xi_{P_j}^{WY}(y, \lambda'|\lambda)$ are the probability distributions representing the X -type and Y -type weak measurements of P_j , respectively. Furthermore, $\xi_{P_j}^{WX}(x|\lambda)$ 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} . Note II in the Supplementary Material provides detailed derivations of these constraints.

Any noncontextual hidden-variable model for the six protocols must satisfy Eqs. (12) to (15). For certain states ρ , this might not be possible, making the realization of the protocols contextual. 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.

- If ρ is such that $\mathcal{N}(\rho) > 3d^2\epsilon$, then these protocols do not admit a noncontextual hidden-variable model.
- If ρ 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 [31, 32]. The original result is written in terms of weak values instead of the KD distribution. In Note III of the Supplementary Material, we provide a proof along similar lines, but in terms of the KD distribution, and with the exact bound on ϵ .

We prove the second part of the theorem by explicitly constructing a noncontextual hidden-variable model that has the same predictions as quantum theory [Eqs. (6) to (8)], and also obeys the Noncontextuality Constraints [Eqs. (12) to (15)]. The set of hidden variables we use is $\Lambda = \{1, 2, \dots, d\}$, and we represent the preparation of ρ 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:

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

$$\xi_{P_j}^{WX}(x, \lambda'|\lambda) = \delta_{\lambda\lambda'} \frac{1}{2}(1 - p_d) + \frac{1}{2}p_d\phi_j(\lambda'|\lambda) + \frac{1}{2}xp_m\delta_{\lambda\lambda'}(2\text{Re}\frac{Q_{j,\lambda'}(\rho)}{\text{Tr}(\Pi_{\lambda'}\rho)} - 1), \quad (17)$$

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

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

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 consider two experimenters, Alice and Bob, each performing an 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 keeping the order of the sequence 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.

Bob lacks knowledge of the order in which Alice has sent her states and thus effectively receives systems in the exotic state ρ_* . From his measurements, Bob obtains the outcome-probability distributions for each measurement procedure [Eqs. (6) to (11)] and determines $Q(\rho_*)$. Since $\mathcal{N}(\rho_*) = 0$, 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 ρ_* [20]. Thus, Bob knows that ρ_* is an exotic state. In Note V of the Supplementary material, we show that Bob can find a $\delta > 0$ such that any pure state decomposition of ρ_* has at least one state ψ_- 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 ψ_- . Bob 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.

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 ρ_* (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.

Furthermore, 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 analogy in entanglement theory [33, 34]. 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 [34]. The converse, however, is not true: The two subsystems can be entangled and yet admit a local hidden-variable model [35, 36]. 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.

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 $\rho_{C,D}$ [33, 34, 37]. The von Neumann entropy $S(\rho_C)$ is a *concave* function in $\rho_{C,D}$. For some states, $S(\rho_C)$ exceeds 0 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. The total KD-nonpositivity $\mathcal{N}(\rho)$ is a *convex* function in ρ . For some (exotic) states ρ_* , $\mathcal{N}(\rho_*)$ equals 0 even if ρ_* 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 $S(\rho_C)$ [34, 38, 39]. Similarly, one can quantify the smallest pure-state-level of KD-nonpositivity of a mixed state ρ via the convex roof of $\mathcal{N}(\rho)$ [40]. 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)

to (8), which requires quantum theory. Future research may strengthen our results to apply to general theories, beyond quantum theory.

Acknowledgments: JJT was supported by the Cambridge Trust. WS was supported by the EPRSC and Hitachi. This work was supported in part by the Agence Nationale de la Recherche under grant 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. SDB thanks Girton College, where part of this work was performed, for its hospitality. The authors thank Christopher K. Long for his helpful comments.

-
- [1] S. Aaronson and D. Gottesman, 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, Rev. Mod. Phys. **84**, 621 (2012).
- [3] V. Veitch, C. Ferrie, D. Gross, and J. Emerson, New Journal of Physics **14**, 113011 (2012).
- [4] V. Veitch, S. A. H. Mousavian, D. Gottesman, and J. Emerson, New Journal of Physics **16**, 013009 (2014).
- [5] R. Raussendorf, J. Bermejo-Vega, E. Tyhurst, C. Okay, and M. Zurek, Phys. Rev. A **101**, 012350 (2020).
- [6] V. Giovannetti, S. Lloyd, and L. Maccone, Phys. Rev. Lett. **96**, 010401 (2006).
- [7] V. Giovannetti, S. Lloyd, and L. Maccone, Nature photonics **5**, 222 (2011).
- [8] L. Maccone, Phys. Rev. A **88**, 042109 (2013).
- [9] D. R. M. Arvidsson-Shukur, N. Yunger Halpern, H. V. Lepage, A. A. Lasek, C. H. W. Barnes, and S. Lloyd, 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. Yunger Halpern, Physical Review Letters **128**, 220504 (2022).
- [11] R. W. Spekkens, Physical Review A **71**, 052108 (2005).
- [12] R. W. Spekkens, Physical review letters **101**, 020401 (2008).
- [13] M. Leifer, Quanta **3**, 67 (2014).
- [14] M. D. Mazurek, M. F. Pusey, R. Kunjwal, K. J. Resch, and R. W. Spekkens, Nature Communications **7**, 10.1038/ncomms11780 (2016).
- [15] D. Schmid and R. W. Spekkens, Phys. Rev. X **8**, 011015 (2018).
- [16] M. Lostaglio, Phys. Rev. Lett. **125**, 230603 (2020).
- [17] J. S. Bell, Physics Physique Fizika **1**, 195 (1964).
- [18] J. G. Kirkwood, Physical Review **44**, 31 (1933).
- [19] P. A. M. Dirac, Rev. Mod. Phys. **17**, 195 (1945).
- [20] D. R. M. Arvidsson-Shukur, W. F. Braasch Jr., S. De Bièvre, J. Dressel, A. N. Jordan, C. Langrenez, M. Lostaglio, J. S. Lundeen, and N. Yunger Halpern, Properties and applications of the kirkwood-dirac distribution (2024), arXiv:2403.18899, arXiv:2403.18899 [quant-ph].
- [21] M. Lostaglio, A. Belenchia, A. Levy, S. Hernández-Gómez, N. Fabbri, and S. Gherardini, Quantum **7**, 1128 (2023).
- [22] J. R. González Alonso, N. Yunger Halpern, and J. Dressel, Phys. Rev. Lett. **122**, 040404 (2019).
- [23] D. Arvidsson-Shukur, J. Chevalier Drori, and N. Yunger Halpern, Journal of Physics A: Mathematical and Theoretical **54**, 284001 (2021).
- [24] S. De Bièvre, Physical Review Letters **127**, 190404 (2021).
- [25] 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 (2024), arXiv:2405.17557 [quant-ph].
- [26] C. Langrenez, D. R. M. Arvidsson-Shukur, and S. De Bièvre, Journal of Mathematical Physics **65**, 072201 (2024).
- [27] Y. Aharonov, D. Z. Albert, and L. Vaidman, Physical review letters **60**, 1351 (1988).
- [28] I. M. Duck, P. M. Stevenson, and E. C. G. Sudarshan, Phys. Rev. D **40**, 2112 (1989).
- [29] R. Jozsa, Phys. Rev. A **76**, 044103 (2007).
- [30] J. Dressel, M. Malik, F. M. Miatto, A. N. Jordan, and R. W. Boyd, Rev. Mod. Phys. **86**, 307 (2014).
- [31] M. F. Pusey, Phys. Rev. Lett. **113**, 200401 (2014).
- [32] R. Kunjwal, M. Lostaglio, and M. F. Pusey, Physical Review A **100**, 042116 (2019).
- [33] V. Vedral, *Introduction to Quantum Information Science* (Oxford University Press, Oxford, England, 2006).
- [34] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Rev. Mod. Phys. **81**, 865 (2009).
- [35] R. F. Werner, Phys. Rev. A **40**, 4277 (1989).
- [36] H. M. Wiseman, S. J. Jones, and A. C. Doherty, Phys. Rev. Lett. **98**, 140402 (2007).
- [37] J. v. Neumann, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse **1927**, 273 (1927).
- [38] S. A. Hill and W. K. Wootters, Phys. Rev. Lett. **78**, 5022 (1997).
- [39] W. K. Wootters, Phys. Rev. Lett. **80**, 2245 (1998).
- [40] C. Langrenez, S. D. Bièvre, and D. R. M. Arvidsson-Shukur, Convex roofs witnessing kirkwood-dirac nonpositivity (2024), arXiv:2407.04558 [quant-ph].

Supplementary Material for Contextuality Can be Verified with Noncontextual Experiments

I. THE QUANTUM THEORETICAL DESCRIPTIONS AND PREDICTIONS OF THE PROTOCOLS

In this note, we describe the protocols in Figure 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 to 3. Last, we show that the outcome distributions of Protocols 4 to 6 match those obtained by marginalizing the outcome distributions of Protocols 2 and 3.

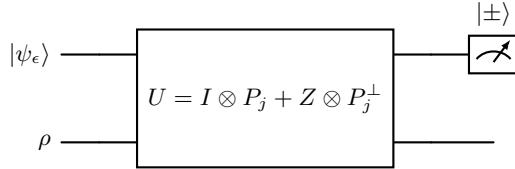


FIG. 1. Weak-Measurement Circuit. This circuit implements the X-type weak measurement on a system in an arbitrary state ρ for the observable P_j . The pointer is initialized in the state $|\psi_\epsilon\rangle := \cos \epsilon |0\rangle + \sin \epsilon |1\rangle$, where ϵ controls the strength of the 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 $y \in \{\pm 1\}$.

A. The Weak Measurement

We consider a weak measurement involving a qubit pointer system [29]. 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 ϵ , which we take to be small. The details are given in Figure 1.

In quantum theory, the weak measurement is represented by Kraus operators. For the X-type weak measurement given in Figure 1, the Kraus operators are

$$N_{x,j} := \langle X = x | U | \psi_\epsilon \rangle = \frac{1}{\sqrt{2}}(\cos \epsilon I + x \sin \epsilon D_j), \quad (\text{S1})$$

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

$$M_{y,j} := \langle Y = y | U | \psi_\epsilon \rangle = \frac{1}{\sqrt{2}}(\cos \epsilon I - iy \sin \epsilon D_j). \quad (\text{S2})$$

B. Protocols 1, 2 and 3

We now use quantum theory to calculate the outcome distributions for Protocols 1 to 3 in Figure 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 Π_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

$$\Pi_k^z = \begin{cases} \Pi_k, & z = +1, \\ I - \Pi_k, & z = -1. \end{cases} \quad (\text{S3})$$

The outcome probability distribution for Protocol 1 is thus

$$f_k^{(1)}(z) = \text{Tr}(\Pi_k^z \rho) := p_z^k. \quad (\text{S4})$$

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. (S1)], and the second measurement corresponds to applying the same projectors as for Protocol 1. We find that

$$\begin{aligned} f_{j,k}^{(2)}(x, z) &= \mathbb{P}(x)\mathbb{P}(z|x), \\ &= \text{Tr}\left(N_{x,j}\rho N_{x,j}^\dagger\right) \text{Tr}\left[\Pi_k^z \frac{N_{x,j}\rho N_{x,j}^\dagger}{\text{Tr}\left(N_{x,j}\rho N_{x,j}^\dagger\right)}\right], \\ &= \text{Tr}\left(\Pi_k^z N_{x,j}\rho N_{x,j}^\dagger\right), \\ &= \frac{1}{2}(1-p_d)p_z^k + \frac{1}{2}xp_m p_z^k[2\text{Re}(w_{j,k}^z) - 1] + \frac{1}{2}p_d p_z^{D_{j,k}}. \end{aligned} \quad (\text{S5})$$

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 $p_z^{D_{j,k}} := \text{Tr}\left(\Pi_k^z D_j \rho D_j^\dagger\right)$ 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

$$w_{j,k}^z = \begin{cases} w_{j,k} & \text{if } z = 1 \\ \frac{\sum_{m \neq k} w_{j,m} p_{+1}^m}{\sum_{m \neq k} p_{+1}^m} & \text{if } z = 0, \end{cases} \quad (\text{S6})$$

where $w_{j,k}$ is the weak-value matrix [27–30] given by

$$w_{j,k} := \frac{\text{Tr}(\Pi_k P_j \rho)}{\text{Tr}(\Pi_k \rho)} = \frac{Q_{j,k}(\rho)}{\text{Tr}(\Pi_k \rho)}. \quad (\text{S7})$$

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. (S2) instead of $N_{x,j}$:

$$f_{j,k}^{(3)}(y, z) = \text{Tr}\left(\Pi_k^z M_{y,j} \rho M_{y,j}^\dagger\right) = \frac{1}{2}(1-p_d)p_z^k + yp_m p_z^k \text{Im}(w_{j,k}^z) + \frac{1}{2}p_d p_z^{D_{j,k}}. \quad (\text{S8})$$

Equations (7) and (8) of the article then follow upon approximating the right-hand side in Eqs. (S5) and (S8) to first order in ϵ and using Eq. (S7) to write the result in terms of KD distributions.

C. Protocols 4, 5 and 6

We now show that, according to quantum theory, the outcome probability distributions for Protocols 4 to 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 Procedure 2 corresponds to discarding the system after performing the X-type weak measurement. We thus effectively implement the following POVM elements by tracing over z :

$$N_{x,j}^\dagger N_{x,j} = (1-p_m)\frac{I}{2} + p_m P_j^x. \quad (\text{S9})$$

Here we introduced 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 ± 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. (S9) exactly corresponds to the measurement process in Protocol 4:

$$\sum_z f_{j,k}^{(2)}(x, z) = f_j^{(4)}(x). \quad (\text{S10})$$

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

$$M_y M_y^\dagger = \frac{1}{2} I. \quad (\text{S11})$$

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

$$\sum_z f_{j,k}^{(3)}(y, z) = f^{(5)}(y). \quad (\text{S12})$$

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 Π_k . Tracing over x corresponds to implementing the quantum channel

$$\rho \mapsto \sum_x N_{x,j} \rho N_{x,j}^\dagger = (1 - p_d) \rho + p_d D_j \rho D_j^\dagger, \quad (\text{S13})$$

D_j is unitary, so this is a well-defined quantum channel. The right-hand side of Eq. (S13) is exactly the transformation process we apply in Protocol 6 before measuring Π_k . We thus find that

$$\sum_x f_{j,k}^{(2)}(x, z) = f_{j,k}^{(6)}(z). \quad (\text{S14})$$

A similar calculation can be used to show that

$$\sum_y f_{j,k}^{(3)}(y, z) = \sum_x f_{j,k}^{(2)}(x, z) = f_{j,k}^{(6)}(z). \quad (\text{S15})$$

II. NONCONTEXTUALITY CONSTRAINTS ESTABLISHED BY PROTOCOLS 1 TO 6

Here, we derive the constraints [Eqs. (12) to (15) of the article] any noncontextual hidden-variable model of Protocols 1 to 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 $\xi_{P_j}^{WX}(x|\lambda)$ for the probability distribution corresponding to discarding the system after an X -type weak measurement. Therefore,

$$\xi_{P_j}^{WX}(x|\lambda) = \int_\Lambda d\lambda' \xi_{P_j}^{WX}(x, \lambda'|\lambda). \quad (\text{S16})$$

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

$$f_j^{(4)}(x) = \sum_z f_{j,k}^{(2)}(x, z). \quad (\text{S17})$$

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

Description	Pictogram	Quantum Model	Hidden-Variable Model
Prepare the state	ρ ——	Density Matrix ρ	$\mu_\rho(\lambda)$
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 $\{N_{x,j}\}_{x \in \{\pm 1\}}$	$\xi_{P_j}^{WX}(x, \lambda' \lambda)$
Weak measurement of P_j using a Y-type pointer		Kraus Operators $\{M_{y,k}\}_{y \in \{\pm 1\}}$	$\xi_{P_j}^{WY}(y, \lambda' \lambda)$
Quantum Channel \mathcal{M}^{D_j}		Quantum Channel $\mathcal{M}^{D_j}(\rho) = D_j \rho D_j^\dagger$ where $D_j = P_j - P_j^\perp$	$\Gamma_{D_j}(\lambda' \lambda)$

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.

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:

$$\xi_{P_j}^{WX}(x|\lambda) = (1 - p_m) \frac{1}{2} + p_m \xi_{P_j}(x|\lambda). \quad (\text{S18})$$

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

$$f_{j,k}^{(6)}(z) = \sum_x f_{j,k}^{(2)}(x, z). \quad (\text{S19})$$

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:

$$\sum_{x \in \{\pm 1\}} \xi_{P_j}^{WX}(x, \lambda'|\lambda) = (1 - p_d) \delta(\lambda - \lambda') + p_d \Gamma_{D_j}(\lambda'|\lambda). \quad (\text{S20})$$

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

(11) in the article:

$$f_{j,k}^{(6)}(z) = \sum_y f_{j,k}^{(3)}(y, z). \quad (\text{S21})$$

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:

$$\sum_{y \in \{\pm 1\}} \xi_{P_j}^{\text{WY}}(y, \lambda' | \lambda) = (1 - p_d)\delta(\lambda - \lambda') + p_d\Gamma_{D_j}(\lambda' | \lambda). \quad (\text{S22})$$

III. KD-NONPOSITIVITY IMPLIES CONTEXTUALITY

In this Note, we prove the first item of Theorem 1 in the article; we show that KD-nonpositivity implies contextuality for sufficiently small weak-measurement strengths ϵ . This result is very closely related to the result in Refs. [31, 32]. 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 ϵ 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 2. *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 ρ 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

$$\mathcal{N}(\rho) = -1 + \sum_{j,k} |Q_{j,k}(\rho)| = \sum_{j,k} (|Q_{j,k}(\rho)| - \text{Re } Q_{j,k}(\rho)). \quad (\text{S23})$$

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

$$\mathcal{N}(\rho) \leq \sum_{j,k} (|\text{Re } Q_{j,k}(\rho)| - \text{Re } Q_{j,k}(\rho) + |\text{Im } Q_{j,k}(\rho)|). \quad (\text{S24})$$

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

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

Lemma 3. *Suppose that for some state ρ , 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 to 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 Figure 2. We will show that for sufficiently small ϵ 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

$$I_\lambda = \{\lambda' \in \Lambda | \xi_{\Pi_k}(+1 | \lambda') \leq \xi_{\Pi_k}(+1 | \lambda)\}, \quad (\text{S25})$$

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

that

$$f_3(+1, +1) = \int_{\Lambda} d\lambda \int_{\Lambda} d\lambda' \mu_{\rho}(\lambda) \xi_{P_j}^{\text{WY}}(+1, \lambda'|\lambda) \xi_{\Pi_k}(+1|\lambda') \quad (\text{S26})$$

$$= \int_{\Lambda} d\lambda \mu_{\rho}(\lambda) \left[\int_{I_{\lambda}} d\lambda' \xi_{P_j}^{\text{WY}}(+1, \lambda'|\lambda) \xi_{\Pi_k}(+1|\lambda') + \int_{\Lambda \setminus I_{\lambda}} d\lambda' \xi_{P_j}^{\text{WY}}(+1, \lambda'|\lambda) \xi_{\Pi_k}(+1|\lambda') \right] \quad (\text{S27})$$

$$\leq \int_{\Lambda} d\lambda \mu_{\rho}(\lambda) \left[\int_{I_{\lambda}} d\lambda' \xi_{P_j}^{\text{WY}}(+1, \lambda'|\lambda) \xi_{\Pi_k}(+1|\lambda) + \int_{\Lambda \setminus I_{\lambda}} d\lambda' \xi_{P_j}^{\text{WY}}(+1, \lambda'|\lambda) \right] \quad (\text{S28})$$

$$\leq \int_{\Lambda} d\lambda \mu_{\rho}(\lambda) \left[\int_{\Lambda} d\lambda' \xi_{P_j}^{\text{WY}}(+1, \lambda'|\lambda) \xi_{\Pi_k}(+1|\lambda) + \int_{\Lambda \setminus I_{\lambda}} d\lambda' \sum_y \xi_{P_j}^{\text{WY}}(y, \lambda'|\lambda) \right]. \quad (\text{S29})$$

In the second line we broke up the integral; in the third line we used the definition of I_{λ} 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. (S29). As $\lambda \notin \Lambda \setminus I_{\lambda}$, the delta function in Eq. (15) of the article vanishes inside the integral. We obtain

$$f_3(+1, +1) \leq \int_{\Lambda} d\lambda \mu_{\rho}(\lambda) \left[\frac{1}{2} \xi_{\Pi_k}(+1|\lambda) + p_d \int_{\Lambda \setminus I_{\lambda}} d\lambda' \Gamma_{D_j}(\lambda'|\lambda) \right] \quad (\text{S30})$$

$$\leq \int_{\Lambda} d\lambda \mu_{\rho}(\lambda) \left[\frac{1}{2} \xi_{\Pi_k}(+1|\lambda) + p_d \right] \quad (\text{S31})$$

$$= \frac{1}{2} p_{+1}^k + p_d, \quad (\text{S32})$$

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 (S8) dictates that it is

$$f_3(+1, +1) = \frac{1}{2} (1 - p_d) p_{+1}^k + p_m p_{+1}^k \text{Im } w_{j,k} + \frac{1}{2} p_z^{D_{j,k}} \quad (\text{S33})$$

$$\geq \frac{1}{2} (1 - p_d) p_{+1}^k + p_m \text{Im } Q_{j,k}(\rho). \quad (\text{S34})$$

Comparing Eqs. (S32) and (S34), we find that

$$\frac{1}{2} (1 - p_d) p_{+1}^k + p_m \text{Im } Q_{j,k}(\rho) \leq \frac{1}{2} p_{+1}^k + p_d. \quad (\text{S35})$$

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

$$\tan \epsilon \geq \frac{4}{2 + p_{+1}^k} \text{Im } Q_{j,k}(\rho) \geq \frac{4}{\pi} \text{Im } Q_{j,k}(\rho), \quad (\text{S36})$$

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

$$\epsilon \geq \text{Im } Q_{j,k}(\rho). \quad (\text{S37})$$

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 to 6 cannot exist for such ϵ .

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. \square

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

Lemma 4. Suppose that for some state ρ , 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 to 6 do not admit a noncontextual hidden-variable model.

Proof. The proof of this lemma is similar to the one of Lemma 3. 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_λ as in the proof of Lemma 3, and through the exact same steps find that

$$f_2(-1, +1) \leq \int_{\Lambda} d\lambda \mu_\rho(\lambda) \left[\int_{\Lambda} d\lambda' \xi_{P_j}^{\text{WX}}(-1, \lambda'|\lambda) \xi_{\Pi_k}(+1|\lambda) + \int_{\Lambda \setminus I_\lambda} d\lambda' \sum_x \xi_{P_j}^{\text{WX}}(x, \lambda'|\lambda) \right]. \quad (\text{S38})$$

Next, we plug Noncontextuality Constraints (13) and (14) from the article into Eq. (S38). Using the same steps as we did for Lemma 3, we find that

$$f_2(-1, +1) \leq \frac{1}{2}(1 + p_m)p_{+1}^k + p_d \quad (\text{S39})$$

Next, we consider the prediction by quantum theory, given by Eq. (S5):

$$f_2(-1, +1) = \frac{1}{2}(1 - p_d)p_{+1}^k - \frac{1}{2}p_m p_{+1}^k(2 \text{Re } w_{j,k} - 1) + \frac{1}{2}p_d p_{+1}^{D_{j,k}} \quad (\text{S40})$$

$$\geq \frac{1}{2}(1 - p_d)p_{+1}^k - p_m \text{Re } Q_{j,k}(\rho) + \frac{1}{2}p_m p_{+1}^k \quad (\text{S41})$$

We now combine Eqs. (S39) and (S41) and use that $p_d = \sin^2 \epsilon$ and that $p_m = \sin 2\epsilon$ to get that

$$\tan \epsilon \geq -\frac{4}{2 + p_{+1}^k} \text{Re } Q_{j,k}(\rho) \geq -\frac{4}{\pi} Q_{j,k}(\rho). \quad (\text{S42})$$

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

$$\epsilon \geq -\text{Re } Q_{j,k}(\rho), \quad (\text{S43})$$

from which the Lemma follows. \square

Combining these Lemmata, we find:

Theorem 5. Let $\delta \in (0, \pi/4]$. If $\mathcal{N}(\rho) > \delta$, then for all $\epsilon < \frac{\delta}{3d^2}$, the Protocols 1 to 6 do not admit a noncontextual hidden-variable model.

IV. KD-POSITIVITY IMPLIES NONCONTEXTUALITY

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

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

We will prove this by explicitly constructing a noncontextual hidden variable model that correctly reproduces the outcomes of Protocols 1 to 6 as predicted by quantum mechanics. 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) to (8) of the article). We thus get the following ‘Correctness Constraints’:

$$\sum_{\lambda \in \Lambda} \mu_\rho(\lambda) \xi_{\Pi_k}(z|\lambda) = f_k^{(1)}(z), \quad (\text{S44})$$

$$\sum_{\lambda, \lambda' \in \Lambda} \mu_\rho(\lambda) \xi_{P_j}^{\text{WX}}(x, \lambda'|\lambda) \xi_{\Pi_k}(z|\lambda') = f_{j,k}^{(2)}(x, z), \quad (\text{S45})$$

$$\sum_{\lambda, \lambda' \in \Lambda} \mu_\rho(\lambda) \xi_{P_j}^{\text{WY}}(y, \lambda'|\lambda) \xi_{\Pi_k}(z|\lambda') = f_{j,k}^{(3)}(y, z), \quad (\text{S46})$$

where we have chosen the ontic space Λ 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) in the article). They are

$$\sum_{\lambda' \in \Lambda} \xi_{P_j}^{\text{WX}}(x, \lambda' | \lambda) = (1 - p_m) \frac{1}{2} + p_m \xi_{P_j}(x | \lambda), \quad (\text{S47})$$

$$\sum_{x \in \{\pm 1\}} \xi_{P_j}^{\text{WX}}(x, \lambda' | \lambda) = (1 - p_d) \delta_{\lambda \lambda'} + p_d \Gamma_{D_j}(\lambda' | \lambda), \quad (\text{S48})$$

$$\sum_{\lambda' \in \Lambda} \xi_{P_j}^{\text{WY}}(y, \lambda' | \lambda) = \frac{1}{2}, \quad (\text{S49})$$

$$\sum_{y \in \{\pm 1\}} \xi_{P_j}^{\text{WY}}(y, \lambda' | \lambda) = (1 - p_d) \delta_{\lambda \lambda'} + p_d \Gamma_{D_j}(\lambda' | \lambda). \quad (\text{S50})$$

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 Γ_{D_j} such that Eqs. (S44) - (S50) 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 Γ_{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 Γ_{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 ξ_{Π_k} to be outcome deterministic:

$$\xi_{\Pi_k}(z | \lambda) = \begin{cases} \delta_{k\lambda} & \text{if } z = +1 \\ 1 - \delta_{k\lambda} & \text{if } z = -1, \end{cases} \quad (\text{S51})$$

which is clearly a normalized probability distribution. Next, we set μ_ρ to be the probability distribution from measuring B . Thus, using the notation introduced in Note I, we set

$$\mu_\rho(\lambda) = p_{+1}^\lambda. \quad (\text{S52})$$

We can now check Eq. (S44) for $z = +1$ by plugging in Eqs. (S51) and (S52):

$$\begin{aligned} \sum_{\lambda \in \Lambda} \mu_\rho(\lambda) \xi_{\Pi_k}(+1 | \lambda) &= \sum_{\lambda \in \mathbb{Z}_d} p_{+1}^\lambda \delta_{k\lambda} \\ &= p_{+1}^k \\ &= f_k^{(1)}(+1), \end{aligned} \quad (\text{S53})$$

which is correct. Furthermore, we may also check (S44) for $z = -1$:

$$\begin{aligned} \sum_{\lambda \in \Lambda} \mu_\rho(\lambda) \xi_{\Pi_k}(-1 | \lambda) &= \sum_{\lambda \in \mathbb{Z}_d} p_{+1}^\lambda (1 - \delta_{k\lambda}) \\ &= 1 - p_{+1}^k \\ &= p_{-1}^k \\ &= f_k^{(1)}(-1), \end{aligned} \quad (\text{S54})$$

which is also as required. Next, to check Eq. (S45), we must provide $\xi_{P_j}^{\text{WX}}$. We use the following ansatz:

$$\xi_{P_j}^{\text{WX}}(x, \lambda' | \lambda) = \delta_{\lambda \lambda'} [\frac{1}{2}(1 - p_d) + \frac{1}{2}xp_m(2 \operatorname{Re} w_{j,\lambda'} - 1)] + \frac{1}{2}p_d \phi_j(\lambda' | \lambda), \quad (\text{S55})$$

where ϕ_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

$$\sum_{\lambda} \phi_j(\lambda' | \lambda) p_{+1}^\lambda = p_{+1}^{D_{j\lambda'}}. \quad (\text{S56})$$

Such a ϕ_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:

$$\begin{aligned} \sum_{\lambda' \in \mathbb{Z}_d, x \in \pm 1} \xi_{P_j}^{\text{WX}}(x, \lambda'|\lambda) &= \sum_{\lambda' \in \mathbb{Z}_d, x \in \pm 1} \left(\delta_{\lambda\lambda'} \left[\frac{1}{2}(1 - p_d) + \frac{1}{2}xp_m(2 \operatorname{Re} w_{j,\lambda'} - 1) \right] + \frac{1}{2}p_d\phi_j(\lambda'|\lambda) \right) \\ &= \sum_{\lambda' \in \mathbb{Z}_d} [\delta_{\lambda\lambda'}(1 - p_d) + p_d\phi_j(\lambda'|\lambda)] \\ &= 1 - p_d + p_d = 1. \end{aligned} \quad (\text{S57})$$

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

Lemma 7. *If $\epsilon < \frac{\sqrt{5}}{5}$, then $\xi_{P_j}^{\text{WX}}(x, \lambda'|\lambda)$ as given in Eq. (S55) 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

$$\begin{aligned} \xi_{P_j}^{\text{WX}}(x, \lambda'|\lambda) &= \delta_{\lambda\lambda'} \left[\frac{1}{2}(1 - p_d) + \frac{1}{2}xp_m(2 \operatorname{Re} w_{j,\lambda'} - 1) \right] + \frac{1}{2}p_d\phi_j(\lambda'|\lambda) \\ &\geq \delta_{\lambda\lambda'} \left[\frac{1}{2}(1 - p_d) + \frac{1}{2}xp_m(2 \operatorname{Re} w_{j,\lambda'} - 1) \right] \\ &\geq \delta_{\lambda\lambda'} \left[\frac{1}{2}\cos^2 \epsilon - \frac{1}{2}\sin 2\epsilon \right] \\ &= \delta_{\lambda\lambda'} h(\epsilon), \end{aligned} \quad (\text{S58})$$

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

$$h'(c)\epsilon + h(0) = h(\epsilon). \quad (\text{S59})$$

There exists some phase ψ such that

$$h'(c) = \frac{1}{2}\sin 2c - \cos 2c = \frac{\sqrt{5}}{2}\sin(2c + \psi) \geq -\frac{\sqrt{5}}{2}. \quad (\text{S60})$$

Substituting (S59) into (S58) and using Eq. (S60) and that $h(0) = \frac{1}{2}$ yields

$$\begin{aligned} \xi_{P_j}^{\text{WX}}(x, \lambda'|\lambda) &= \delta_{\lambda\lambda'}[h'(c)\epsilon + h(0)] \\ &\geq \delta_{\lambda\lambda'} \left[\frac{-\sqrt{5}}{2}\epsilon + \frac{1}{2} \right] \\ &\geq 0 \end{aligned} \quad (\text{S61})$$

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

We have thus found that the ansatz (S55) is a normalized probability distribution. We proceed to check that it indeed satisfies Eq. (S45) for $z = +1$.

$$\begin{aligned} \sum_{\lambda, \lambda' \in \Lambda} \mu_\rho(\lambda) \xi_{P_j}^{\text{WX}}(x, \lambda'|\lambda) \xi_{\Pi_k}(+1|\lambda') &= \sum_{\lambda, \lambda' \in \mathbb{Z}_d} p_{+1}^\lambda \xi_{P_j}^{\text{WX}}(x, \lambda'|\lambda) \delta_{k\lambda'} \\ &= \sum_{\lambda \in \mathbb{Z}_d} p_{+1}^\lambda \xi_{P_j}^{\text{WX}}(x, k|\lambda) \\ &= \sum_{\lambda \in \mathbb{Z}_d} p_{+1}^\lambda \left\{ \delta_{\lambda k} \left[\frac{1}{2}(1 - p_d) + \frac{1}{2}xp_m(2 \operatorname{Re} w_{j,k} - 1) \right] + \frac{1}{2}p_d\phi_j(k|\lambda) \right\} \\ &= \frac{1}{2}(1 - p_d)p_{+1}^k + \frac{1}{2}xp_m p_{+1}^k(2 \operatorname{Re} w_{j,k} - 1) + \frac{1}{2}p_d \sum_{\lambda \in \mathbb{Z}_d} \phi_j(k|\lambda) p_{+1}^\lambda \\ &= \frac{1}{2}(1 - p_d)p_{+1}^k + \frac{1}{2}xp_m p_{+1}^k(2 \operatorname{Re} w_{j,k} - 1) + \frac{1}{2}p_d p_{+1}^{D_{j,k}} \\ &= f_{j,k}^{(2)}(x, +1), \end{aligned} \quad (\text{S62})$$

which is correct. Similarly, we may check Eq. (S45) for $z = -1$:

$$\begin{aligned}
\sum_{\lambda, \lambda' \in \Lambda} \mu_\rho(\lambda) \xi_{P_j}^{\text{WX}}(x, \lambda' | \lambda) \xi_{\Pi_k}(-1 | \lambda') &= \sum_{\lambda, \lambda' \in \mathbb{Z}_d} p_S^\lambda \xi_{P_j}^{\text{WX}}(x, \lambda' | \lambda) (1 - \delta_{k\lambda'}) \\
&= \sum_{\lambda \in \mathbb{Z}_d, \lambda' \in \mathbb{Z}_d \setminus \{k\}} p_S^\lambda \xi_{P_j}^{\text{WX}}(x, \lambda' | \lambda) \\
&= \sum_{\lambda' \in \mathbb{Z}_d \setminus \{k\}} \frac{1}{2}(1 - p_d)p_S^{\lambda'} + \frac{1}{2}xp_m p_S^{\lambda'} (2 \operatorname{Re} w_{j\lambda'} - 1) + \frac{1}{2}p_d p_S^{D_j \lambda'} \\
&= \frac{1}{2}(1 - p_d)p_{-1}^k + \frac{1}{2}xp_m p_{-1}^k \left(2 \frac{\sum_{\lambda' \in \mathbb{Z}_d \setminus \{k\}} p_S^{\lambda'} \operatorname{Re} w_{j\lambda'}}{p_{-1}^k} - 1 \right) + \frac{1}{2}p_d p_{-1}^{D_j k} \\
&= \frac{1}{2}(1 - p_d)p_{-1}^k + \frac{1}{2}xp_m p_{-1}^k (2g_{-1}(\operatorname{Re} w_{jk}) - 1) + \frac{1}{2}p_d p_{-1}^{D_j k} \\
&= f_{j,k}^{(2)}(x, -1),
\end{aligned} \tag{S63}$$

which is also correct. We thus conclude that our ansatz satisfies Eq. (S45).

We move on to checking the Noncontextuality Constraints involving $\xi_{P_j}^{\text{WX}}$. We start with Eq. (S47) for $x = +1$.

$$\begin{aligned}
\sum_{\lambda' \in \Lambda} \xi_{P_j}^{\text{WX}}(+1, \lambda' | \lambda) &= \sum_{\lambda' \in \mathbb{Z}_d} \left(\delta_{\lambda\lambda'} \left[\frac{1}{2}(1 - p_d) + \frac{1}{2}p_m (2 \operatorname{Re} w_{j,\lambda'} - 1) \right] + \frac{1}{2}p_d \phi_j(\lambda' | \lambda) \right) \\
&= \frac{1}{2}(1 - p_d) + \frac{1}{2}p_m (2 \operatorname{Re} w_{j,\lambda} - 1) + \frac{1}{2}p_d \\
&= (1 - p_m) \frac{1}{2} + p_m \operatorname{Re} w_{j,\lambda}
\end{aligned} \tag{S64}$$

Comparing this with the right-hand side of Eq. (S47) for $x = +1$, we see that we must set $\xi_{P_j}(+1 | \lambda) = \operatorname{Re} w_{j,\lambda}$. This yields probabilities in $[0, 1]$ since the fact that ρ is KD-positive implies that the weak values lie in $[0, 1]$ by Eq. (S7). Furthermore, we fix the $\xi_{P_j}(+1 | \lambda)$ via normalization: $\xi_{P_j}(-1 | \lambda) = 1 - \operatorname{Re} w_{j,\lambda}$. We can now also check Eq. (S47) for $x = -1$:

$$\begin{aligned}
\sum_{\lambda' \in \Lambda} \xi_{P_j}^{\text{WX}}(-1, \lambda' | \lambda) &= \sum_{\lambda' \in \mathbb{Z}_d} \left(\delta_{\lambda\lambda'} \left[\frac{1}{2}(1 - p_d) - \frac{1}{2}p_m (2 \operatorname{Re} w_{j,\lambda'} - 1) \right] + \frac{1}{2}p_d \phi_j(\lambda' | \lambda) \right) \\
&= \frac{1}{2}(1 - p_d) - \frac{1}{2}p_m (2 \operatorname{Re} w_{j,\lambda} - 1) + \frac{1}{2}p_d \\
&= (1 - p_m) \frac{1}{2} - p_m (1 - \operatorname{Re} w_{j,\lambda}) \\
&= (1 - p_m) \frac{1}{2} - p_m \xi_{P_j}(-1 | \lambda)
\end{aligned} \tag{S65}$$

as required. We thus conclude that Noncontextuality Constraint (S47) is satisfied. We move on to showing that Noncontextuality Constraint (S48) is satisfied. We calculate its LHS for this ansatz:

$$\begin{aligned}
\sum_{x \in \{\pm 1\}} \xi_{P_j}^{\text{WX}}(x, \lambda' | \lambda) &= \sum_{x \in \{\pm 1\}} \left(\delta_{\lambda\lambda'} \left[\frac{1}{2}(1 - p_d) + \frac{1}{2}xp_m (2 \operatorname{Re} w_{j,\lambda'} - 1) \right] + \frac{1}{2}p_d \phi_j(\lambda' | \lambda) \right) \\
&= (1 - p_d)\delta_{\lambda\lambda'} + p_d \phi_j(\lambda' | \lambda).
\end{aligned} \tag{S66}$$

Comparing this result to the right-hand side of Eq. (S48), 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. (S48) 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:

$$\xi_{P_j}^{\text{WY}}(y, \lambda' | \lambda) = \delta_{\lambda\lambda'} \left[\frac{1}{2}(1 - p_d) + yp_m \operatorname{Im} w_{j,\lambda'} \right] + \frac{1}{2}p_d \phi_j(\lambda' | \lambda), \tag{S67}$$

where ϕ_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 will omit it here. First, we check Eq. (S46) for $z = +1$:

$$\begin{aligned}
\sum_{\lambda, \lambda' \in \Lambda} \mu_\rho(\lambda) \xi_{P_j}^{\text{WY}}(y, \lambda' | \lambda) \xi_{\Pi_k}(+1 | \lambda') &= \sum_{\lambda, \lambda' \in \mathbb{Z}_d} p_{+1}^\lambda \xi_{P_j}^{\text{WY}}(y, \lambda' | \lambda) \delta_{\lambda' k} \\
&= \sum_{\lambda \in \mathbb{Z}_d} p_{+1}^\lambda \xi_{P_j}^{\text{WY}}(y, k | \lambda) \\
&= \sum_{\lambda \in \mathbb{Z}_d} \left(p_{+1}^\lambda \delta_{\lambda k} \left[\frac{1}{2}(1 - p_d) + y p_m \operatorname{Im} w_{j,k} \right] + \frac{1}{2} p_d \phi_j(k | \lambda) \right) \\
&= \frac{1}{2}(1 - p_d) p_{+1}^k + y p_m p_{+1}^k \operatorname{Im} w_{j,k} + \frac{1}{2} p_d \sum_{\lambda \in \mathbb{Z}_d} p_{+1}^\lambda \phi_j(k | \lambda) \\
&= \frac{1}{2}(1 - p_d) p_{+1}^k + y p_m p_{+1}^k \operatorname{Im} w_{j,k} + \frac{1}{2} p_d p_{+1}^{D_{j,k}} \\
&= f_{j,k}^{(3)}(y, +1).
\end{aligned} \tag{S68}$$

Hence Eq. (S46) is satisfied for $z = +1$. We also check Eq. (S49) for $z = -1$:

$$\begin{aligned}
\sum_{\lambda, \lambda' \in \Lambda} \mu_\rho(\lambda) \xi_{P_j}^{\text{WY}}(y, \lambda' | \lambda) \xi_{\Pi_k}(-1 | \lambda') &= \sum_{\lambda, \lambda' \in \mathbb{Z}_d} p_{+1}^\lambda \xi_{P_j}^{\text{WY}}(y, \lambda' | \lambda) (1 - \delta_{\lambda' k}) \\
&= \sum_{\lambda \in \mathbb{Z}_d, \lambda' \in \mathbb{Z}_d \setminus \{k\}} p_{+1}^\lambda \xi_{P_j}^{\text{WY}}(y, \lambda' | \lambda) \\
&= \sum_{\lambda' \in \mathbb{Z}_d \setminus \{k\}} \left(\frac{1}{2}(1 - p_d) p_{+1}^{\lambda'} + y p_m p_{+1}^{\lambda'} \operatorname{Im} w_{j,\lambda'} + \frac{1}{2} p_d p_{+1}^{D_{j,\lambda'}} \right) \\
&= \frac{1}{2}(1 - p_d) p_{-1}^k + y p_m p_{-1}^k \frac{\sum_{\lambda' \in \mathbb{Z}_d \setminus \{k\}} p_{+1}^{\lambda'} \operatorname{Im} w_{j,\lambda'}}{p_{-1}^k} + \frac{1}{2} p_d p_{-1}^{D_{j,k}} \\
&= \frac{1}{2}(1 - p_d) p_{-1}^k + y p_m p_{-1}^k g_{-1}(\operatorname{Im} w_{j,\lambda'}) + \frac{1}{2} p_d p_{-1}^{D_{j,k}} \\
&= f_{j,k}^{(3)}(y, -1).
\end{aligned} \tag{S69}$$

Hence Eq. (S49) 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. (S49).

$$\begin{aligned}
\sum_{\lambda' \in \Lambda} \xi_{P_j}^{\text{WY}}(y, \lambda' | \lambda) &= \sum_{\lambda' \in \mathbb{Z}_d} \left(\delta_{\lambda \lambda'} \left[\frac{1}{2}(1 - p_d) + y p_m \operatorname{Im} w_{j,\lambda'} \right] + \frac{1}{2} p_d \phi_j(\lambda' | \lambda) \right) \\
&= \sum_{\lambda' \in \mathbb{Z}_d} \delta_{\lambda \lambda'} \frac{1}{2}(1 - p_d) + \frac{1}{2} p_d \phi_j(\lambda' | \lambda) \\
&= \frac{1}{2}(1 - p_d) + \frac{1}{2} p_d = \frac{1}{2},
\end{aligned} \tag{S70}$$

where to go from the first to the second line we used that by Eq. (S7) the fact that ρ is KD-positive implies that $\operatorname{Im} w_{j,\lambda'} = 0$. Hence Eq. (S49) is satisfied. Last, we need to check Eq. (S50).

$$\begin{aligned}
\sum_{y \in \{\pm 1\}} \xi_{P_j}^{\text{WY}}(y, \lambda' | \lambda) &= \sum_{y \in \{\pm 1\}} \left(\delta_{\lambda \lambda'} \left[\frac{1}{2}(1 - p_d) + y p_m \operatorname{Im} w_{j,\lambda'} \right] + \frac{1}{2} p_d \phi_j(\lambda' | \lambda) \right) \\
&= (1 - p_d) \delta_{\lambda \lambda'} + p_d \phi_j(\lambda' | \lambda) \\
&= (1 - p_d) \delta_{\lambda \lambda'} + p_d \Gamma_{D_j}(\lambda' | \lambda)
\end{aligned} \tag{S71}$$

where we used that we chose ϕ_j earlier to be Γ_{D_j} . Hence we have explicitly demonstrated that if the state is KD-positive and ϵ is sufficiently small, that then there exists a noncontextual hidden variable model for Protocols 1 to 6.

V. 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 Note. More precisely, we prove the following statement:

Lemma 8 (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 ρ_\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 9 (Properties of \mathcal{A}). *The set \mathcal{A} has the following properties*

- (i) *For every decomposition of an exotic state ρ_\star into pure states, $\rho_\star = \sum_i p_i \psi_i$, there exists at least one pure state ψ_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. (i) by contradiction. Suppose that for all i , $\psi_i \notin \mathcal{A}$. We then calculate

$$\text{Tr}(H\rho_\star) = \sum_i p_i \text{Tr}(H\psi_i) < \sum_i p_i d = d, \quad (\text{S72})$$

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.

(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 ρ_{\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 , we this yields

$$\text{Tr}(H\rho_{\min}) \leq c < d. \quad (\text{S73})$$

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. \square

The proof of Lemma 8 then straightforwardly follows from Lemma 9.