White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/152481/

Version: Published Version

Article:
Kunjwal, Ravi, Lostaglio, Matteo and Pusey, Matthew Fairbairn (2019) Anomalous weak values and contextuality: Robustness, tightness, and imaginary parts. Physical Review A. 042116. ISSN: 1094-1622

https://doi.org/10.1103/PhysRevA.100.042116

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Anomalous weak values and contextuality: Robustness, tightness, and imaginary parts

Ravi Kunjwal,[1] Matteo Lostaglio,[2] and Matthew F. Pusey[3]

[1]*Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo, Ontario, Canada N2L 2Y5*

[2]*ICFO–Institut de Ciencies Fotoniques, Barcelona Institute of Science and Technology, 08860 Castelldefels (Barcelona), Spain*

[3]*Department of Computer Science, University of Oxford, Wolfson Building, Parks Road, Oxford OX1 3QD, United Kingdom*

(Received 4 September 2019; published 21 October 2019)

Weak values are quantities accessed through quantum experiments involving weak measurements and postselection. It has been shown that "anomalous" weak values (those lying beyond the eigenvalue range of the corresponding operator) defy classical explanation in the sense of requiring contextuality [M. F. Pusey, Phys. Rev. Lett. **113**, 200401 (2014)]. Here we elaborate on and extend that result in several directions. First, the original theorem requires certain perfect correlations that can never be realized in any actual experiment. Hence, we provide theorems that allow for a noise-robust experimental verification of contextuality from anomalous weak values, and compare with a recent experiment. Second, the original theorem connects the anomaly to contextuality only in the presence of a whole set of extra operational constraints. Here we clarify the debate surrounding anomalous weak values by showing that these conditions are tight: if any one of them is dropped, the anomaly can be reproduced classically. Third, whereas the original result required the real part of the weak value to be anomalous, we also give a version for any weak value with nonzero imaginary part. Finally, we show that similar results hold if the weak measurement is performed through qubit pointers, rather than the traditional continuous system. In summary, we provide inequalities for witnessing nonclassicality using experimentally realistic measurements of any anomalous weak value, and clarify what ingredients of the quantum experiment must be missing in any classical model that can reproduce the anomaly.

DOI: 10.1103/PhysRevA.100.042116

I. INTRODUCTION

Weak measurements [1] are a class of minimally disturbing quantum measurements whose practical as well as foundational relevance is currently being investigated [2]. A weak measurement of an observable O can be realized by weakly coupling a quantum system to a one-dimensional pointer device via a von Neumann–type interaction $\propto O \otimes \Gamma$, with Γ the momentum of the pointer, so that a small amount of information is imprinted in the pointer at the cost of a small disturbance on the system.

Pivotal to any attempt to establish the presence of nonclassical effects in a given experiment is the formulation of a rigorous no-go theorem based on a precise and operational notion of nonclassicality. It has long been argued that the average final position of the pointer—conditioned upon a successful postselection performed on the system after the weak measurement—is a witness to nonclassicality [1]; in the quantum formalism this quantity is related to the (real part of the) *weak value*, which is ${}_{\phi}\langle O\rangle_{\psi} := \langle\phi|O|\psi\rangle / \langle\phi|\psi\rangle$, where O is the observable being weakly measured, $|\psi\rangle$ is the initial preparation, and $|\phi\rangle$ is the postselection. A long-standing debate ensued between those supporting the thesis that these experiments are indeed probing truly quantum effects and those arguing that they can be easily understood from classical statistics [3–8].

Recently, a precise no-go theorem was established [9]. The theorem proves that *anomalous weak values* (AWVs), i.e., ${}_{\phi}\langle O\rangle_{\psi}$ taking values beyond the spectrum of O, are associated with operational statistics defying any noncontextual explanation in the generalized sense introduced by Spekkens [10]. Nevertheless, the theorem of Ref. [9] leaves several questions open:

(1) First of all, it assumes an *exactly* projective postselection $|\phi\rangle$, which makes any experimental test [11] necessarily inconclusive; in fact, any degree of noise makes the no-go theorem inapplicable. Does the nonclassicality of AWVs survive real-world conditions?

(2) Second, both Ref. [9] and the noise-robust theorems presented here prove that AWVs are nonclassical in the presence of a set of extra operational conditions. Are these all truly necessary?

(3) Third, the theorem only refers to the *real part* of the weak value. Is a nonzero value of the imaginary part of the weak value also nonclassical?

(4) Fourth, the relation between AWVs and contextuality holds for a measurement with a continuum of outcomes. Can it be extended to discrete systems, such as an experiment involving only a single qubit pointer, or a coarse graining of the standard weak-value experiment? This is also experimentally relevant because the infinitely many operational constraints required for the original theorem [9] to hold cannot be tested by finite means, and a discrete pointer is often more practical anyway.

(5) Finally, the theorem identifies a single noncontextuality inequality which is violated in the presence of AWVs. However, is the inequality unique and is it tight?

 ©2019 American Physical Society

Our investigation largely answers all these questions:

(1) We provide two proofs of contextuality from AWVs that are robust to noise, based on Theorems 1 and 4. The two proofs are complementary, each requiring the satisfaction of a different set of operational constraints together with the observation of the AWVs. These results show that at the price of extending the set of operational tests required, the relation between AWVs and nonclassicality extends beyond the ideal, noiseless case. We also discuss the significance of these results for current experimental tests (Sec. III B).

(2) We show that the extra operational conditions in our theorems form a minimal set: dropping any one of them allows us to reproduce the AWVs within a classical model (Sec. V). This illuminates the debate around "quantumness" of AWVs (e.g., [5–7]), since it rigorously shows that it is only in the presence of *all* the operational facts listed in our theorems that AWVs defy a classical explanation.

(3) The imaginary part of the weak value admits its own contextuality theorem (Sec. II E, Theorem 2). Hence, *any* AWVs can be related to contextuality. We clarify why this is not in contradiction with recent studies [12,13] suggesting that imaginary weak values admit a classical model.

(4) The contextuality of AWVs has nothing to do with continuous measurements and extends to discrete pointers as well (Sec. II F, Theorem 3). This makes the experiment suited for conclusive experimental verification, since in this case only a finite set of operational tests is required.

(5) The noncontextual bound in Ref. [9] is not tight, but we provide an improved version and investigate its tightness and uniqueness using computational methods from Ref. [14] (Appendix B).

Our theorems are noise-robust in the sense of not requiring perfectly projective measurements, but noise can also impact the other operational conditions of our proofs. We view these issues as being outside the scope of this work, because with the form of noise robustness we provide in place, there are generic approaches to tackling the main remaining idealizations, as discussed in Sec. IV.

II. NOISE-ROBUST NO-GO THEOREMS FOR ANOMALOUS WEAK VALUES

A. Weak values

Let ρ be a quantum state, O an observable, and $[y|M_F]$ a postselection measurement, i.e., $[y=1|M_F]=\Pi_\phi$ (successful postselection), $[y=0|M_F]=\mathbb{1}-\Pi_\phi$ (failed postselection), with $\Pi_\phi=|\phi\rangle\langle\phi|$. We can then define the (generalized) *weak value*

$$ {}_\phi\langle O\rangle_\rho=\frac{\mathrm{Tr}(\Pi_\phi O\rho)}{\mathrm{Tr}(\Pi_\phi\rho)}. \tag{1}$$

This expression equals the standard expression of the weak value of Ref. [1] when $\rho=|\psi\rangle\langle\psi|$. For ${}_\phi\langle O\rangle_\rho$ to be well defined, we take $\mathrm{Tr}\,(\Pi_\phi\rho)>0$; i.e., the preselection and postselection are nonorthogonal. The weak value can be experimentally accessed by a weak measurement of O. Specifically, couple O with a one-dimensional pointer device through the Hamiltonian $H=O\otimes\Gamma$, with Γ the momentum of the pointer. Suppose the pointer is initialized in a Gaussian pure state centered around the origin and with spread s:

$$|\psi\rangle_P=\int dx G_s(x)\,|x\rangle\,,$$
$$G_s(x)=(\pi s^2)^{-1/4}\exp[-x^2/(2s^2)]. \tag{2}$$

In the limit $s\to\infty$, if a projective measurement of the pointer's position is carried out after a unit time, we obtain a so-called *weak measurement* of O ($s\to 0$ would give a projective measurement of O).

Suppose now the postselection measurement $\{\Pi_\phi,\mathbb{1}-\Pi_\phi\}$ is carried out on the system, after the interaction with the pointer. The average position of the pointer, conditioned on observing Π_ϕ (successful postselection), is proportional to $\mathrm{Re}({}_\phi\langle O\rangle_\rho)$, whereas $\mathrm{Im}({}_\phi\langle O\rangle_\rho)$ can be recovered from the expected momentum of the pointer given a successful postselection [15].

The weak value is called *anomalous* when it cannot be written as a convex combination of the eigenvalues of O. There are two ways this can happen:

(i) $\mathrm{Re}({}_\phi\langle O\rangle_\rho)$ is smaller than the smallest eigenvalue of O, or larger than the largest eigenvalue;

(ii) $\mathrm{Im}({}_\phi\langle O\rangle_\rho)\neq 0$.

Only (i) was related to contextuality in Ref. [9], but our results here show that both in fact lead to proofs of contextuality.

Writing the spectral decomposition of O as $O=\sum_i o_i\mathcal{E}_i$, we have that

$${}_\phi\langle O\rangle_\rho=\sum_i o_i\,{}_\phi\langle\mathcal{E}_i\rangle_\rho \tag{3}$$

and $\sum_i {}_\phi\langle\mathcal{E}_i\rangle_\rho={}_\phi\langle\mathbb{1}\rangle_\rho=1$. Then, if ${}_\phi\langle O\rangle_\rho$ is anomalous, at least one of the ${}_\phi\langle\mathcal{E}_i\rangle_\rho$ must be anomalous (i.e., not a standard probability).[1] This is because if all the ${}_\phi\langle\mathcal{E}_i\rangle_\rho$ are standard probabilities then (3) shows that ${}_\phi\langle O\rangle_\rho$ is in the convex hull of the o_i.

Since, then, whenever we have an anomalous weak value for an observable O we can also find an anomalous weak value for one of its eigenprojectors, without loss of generality we will focus on weak values of projectors.

Furthermore, if a projector $\mathcal{E}$ is anomalous due to its real part, then either $\mathrm{Re}_\phi\langle\mathcal{E}\rangle_\rho<0$ or $\mathrm{Re}_\phi\langle(\mathbb{1}-\mathcal{E})\rangle_\rho<0$; similarly, if a projector $\mathcal{E}$ is anomalous due to its imaginary part, then either $\mathrm{Im}_\phi\langle\mathcal{E}\rangle_\rho<0$ or $\mathrm{Im}_\phi\langle(\mathbb{1}-\mathcal{E})\rangle_\rho<0$. Hence, without loss of generality we will focus on anomalous weak values for projectors with negative real or imaginary part.

For calculations it will often be useful to refer to the numerator of Eq. (1), which we write as $\langle\Pi_\phi\mathcal{E}\rangle_\rho:=\mathrm{Tr}(\Pi_\phi\mathcal{E}\rho)$.[2]

[1] Note that one can have instances in which some or all ${}_\phi\langle\mathcal{E}_i\rangle_\rho$ are anomalous, but ${}_\phi\langle O\rangle_\rho$ is not; e.g., if an observable has a zero eigenvalue then the weak value of the associated projector is irrelevant to the weak value of the observable.

[2] $\langle\Pi_\phi\mathcal{E}\rangle_\rho$ coincides with the so-called Kirkwood-Dirac [16,17] quasiprobability distribution, the real part of which is the Margenau-Hills [18] distribution; see Sec. IV A of Ref. [19] for details. These distributions are related to the "optimal" estimate of the observable $\mathcal{E}$ from a measurement of Π_ϕ, under the prior information that the initial state is ρ [20].

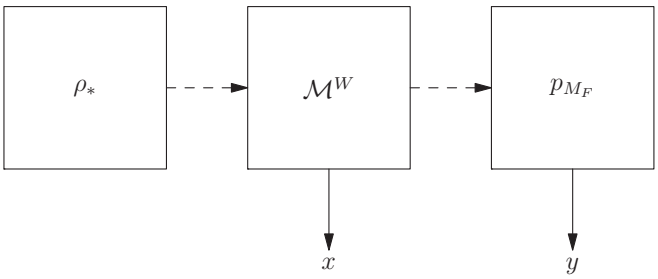

FIG. 1. Illustration of the three stages of a quantum weak-value experiment.

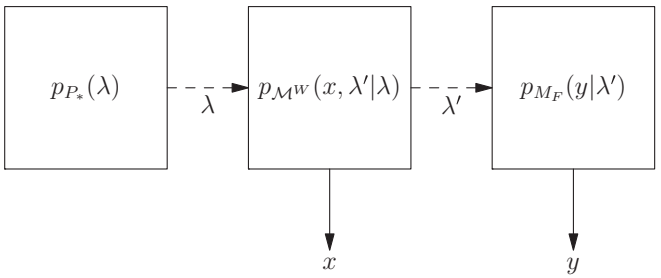

FIG. 2. Illustration of an ontological model for the quantum experiment in Fig. 1.

Since the denominator $\mathrm{Tr}(\Pi_\phi\rho)$ is a positive real number (in particular, recalling that it must be nonzero for a well-defined weak value), $\langle\Pi_\phi\mathcal{E}\rangle_\rho$ has negative real or imaginary parts if and only if ${}_\phi\langle\mathcal{E}\rangle_\rho$ does.

B. Setting the stage: The standard quantum experiment

Let us discuss the traditional experimental setting for weak measurements and weak values [1] (see Appendix A for some details of the calculations; later we will discuss extensions to qubit pointers). As discussed above, we can focus on the weak value of some projector $\mathcal{E}$. There are three stages of the quantum experiment (see Fig. 1):

Preparation. A system is prepared in some quantum state. Since no difficulties arise from allowing a generic mixed state ρ_*, we allow mixed preparations.

Weak measurement. A measurement is performed through the following scheme: a pointer device, represented by a one-dimensional continuous system with conjugate variables X and Γ, is initialized in the Gaussian pure state $|\psi\rangle_P$ given above. The system is coupled to the pointer through the Hamiltonian $H = \mathcal{E}\otimes\Gamma$.

A standard calculation (see, e.g., the proof of Theorem 1 in Ref. [9]) shows that after a unit time, a measurement of X on the pointer realizes a positive operator valued measure (POVM) $[x|M_W] = N_x^\dagger N_x$ on the system given by

$$N_x = \langle x|e^{-iH}|\Psi\rangle_P = G_s(x-1)\mathcal{E} + G_s(x)\mathcal{E}^\perp, \tag{4}$$

$$[x|M_W] = G_s^2(x-1)[y=1|M_\mathcal{E}] + G_s^2(x)[y=0|M_\mathcal{E}], \tag{5}$$

where $[y=1|M_\mathcal{E}] = \mathcal{E}$, $[y=0|M_\mathcal{E}] = \mathcal{E}^\perp = 1-\mathcal{E}$.

Let $\mathcal{M}_x^W(\cdot) = N_x(\cdot)N_x^\dagger$ be the state update map for outcome x. The channel induced by the weak measurement when the outcome is not recorded is

$$\mathcal{M}(\cdot) = \int_{-\infty}^{+\infty} dx\mathcal{M}_x^W(\cdot) = \int_{-\infty}^{+\infty} dx N_x(\cdot)N_x^\dagger. \tag{6}$$

One finds $\mathcal{M}(\rho) = (1-p_d)\rho + p_d(\mathcal{E}-\mathcal{E}^\perp)\rho(\mathcal{E}-\mathcal{E}^\perp)$, with a "probability of disturbance" $p_d = \frac{1-e^{-1/4s^2}}{2}$. Hence,

$$\mathcal{M} = (1-p_d)\mathcal{I} + p_d\mathcal{M}^D, \tag{7}$$

with $\mathcal{M}^D(\rho) := (\mathcal{E}-\mathcal{E}^\perp)\rho(\mathcal{E}-\mathcal{E}^\perp)$.

Postselection. Finally, one can measure $[y|M_F]$ and compute the probability of a negative x followed by a successful postselection

$$p_-^{\text{ideal}} = \int_{-\infty}^{0} dx\,\mathrm{Tr}(\Pi_\phi N_x\rho_* N_x^\dagger) = \int_{-\infty}^{0} dx\mathrm{Tr}\big[\Pi_\phi\mathcal{M}_x^W(\rho_*)\big],$$

which will be a central witness of nonclassicality in the following theorems. Denoting the undisturbed probability of postselection by $p_F = \mathrm{Tr}\,(\Pi_\phi\rho_*)$, one finds

$$p_-^{\text{ideal}} = \frac{p_F}{2} - \frac{\mathrm{Re}(\langle\Pi_\phi\mathcal{E}\rangle_{\rho_*})}{\sqrt{\pi}s} + o\left(\frac{1}{s}\right). \tag{8}$$

This is a simple calculation; see, e.g., the proof of Lemma 1 in Ref. [21] (note, however, that we redefined p_-^{ideal} without the normalization by the postselection probability). Recall from the previous section that a weak value with an anomalous real part implies an $\mathcal{E}$ with $\mathrm{Re}(\langle\Pi_\phi\mathcal{E}\rangle_{\rho_*}) < 0$. We will show that this means p_-^{ideal} is larger than can be explained in a noncontextual model.

C. Noncontextual description of the quantum experiment

We now analyze how a putative noncontextual ontological model (Fig. 2) would describe the quantum experiment (Fig. 1). Let us follow the three stages:

Preparation. The preparation of the quantum state ρ_* can be abstractly thought of as a set of instructions P_* that initialize the system. In an ontological model, this is associated with sampling from a distribution $p_{P_*}(\lambda)$ over some set of hidden variables λ.

Weak measurement. The weak measurement is a quantum instrument $\{\mathcal{M}_x^W\}$ (also understood as a set of experimental procedures) that, in the ontological model, is represented by the function $p_{\mathcal{M}_W}(x,\lambda'|\lambda)$. This describes the probability that given as input the state λ, the weak measurement gives outcome x and updates the state to λ' (the update $\lambda\to\lambda'$ models the potential disturbance induced by the measuring apparatus). If $p_\mathcal{M}(\lambda'|\lambda)$ represents the matrix of transition probabilities associated to the channel $\mathcal{M}$ in Eq. (6), one has $p_\mathcal{M}(\lambda'|\lambda) = \int_{-\infty}^{+\infty} dx p_{\mathcal{M}_W}(x,\lambda'|\lambda)$. On the other hand, the response function $p_{M_W}(x|\lambda)$ of the weak measurement $[x|M_W]$, giving the probability that the weak measurement outputs x given the input state λ, is given by $p_{M_W}(x|\lambda) = \int d\lambda' p_{\mathcal{M}_W}(x,\lambda'|\lambda)$.

Postselection. The measurement $[y|M_F]$ is also represented in the ontological model by a response function $p_{\mathcal{M}_F}(y|\lambda)$. While in the quantum experiment $[y|M_F]$ would ideally be a projective measurement, in contrast to Ref. [9], our theorems

will not rely on this being the case (in fact, our first theorem makes no assumption about $[y|M_F]$). This is necessary in any experimental verification of the relation between anomalous weak values and contextuality, since no experiment can achieve this idealization.

Operational statistics. The operational statistics collected by the whole experiment is summarized by the following:

(1) $p(x, y|P_*, \mathcal{M}^W, M_F)$, the probability that if the preparation procedure P_* is followed, sequentially performing the weak measurement procedure $\mathcal{M}^W$ and the postselection procedure M_F returns outcomes x and y, respectively. In the quantum setting this is given by $\mathrm{Tr}\{[y|M_F]\mathcal{M}^W_x(\rho_*)\}$.

(2) $p(y|P_*, M_F)$, the probability that if the preparation procedure P_* is directly followed by the postselection measurement procedure M_F, one gets outcome y. In the quantum setting this is given by $\mathrm{Tr}\,([y|M_F]\rho_*)$.

An ontological model for this experiment is a set of assignments as described above and satisfying

$$p(x, y|P_*, \mathcal{M}^W, M_F) = \int d\lambda' d\lambda p_{P_*}(\lambda) p_{\mathcal{M}_W}(x, \lambda'|\lambda) p_{M_F}(y|\lambda'),$$

$$p(y|P_*, M_F) = \int d\lambda p_{P_*}(\lambda) p_{M_F}(y|\lambda).$$

Noncontextuality. A generic ontological model description of the experiment can always be found, whatever the operational statistics. However, noncontextual models (according to the definition of Ref. [10]) are those that associate with operationally indistinguishable procedures identical representation in the ontological model. In the present case, the weak measurement procedure $[x|M_W]$ is operationally equivalent, due to Eq. (5), to measuring $[y|M_\mathcal{E}]$ and then sampling as prescribed according to the distribution $G^2_s(x)$. Hence noncontextual models require

$$p_{M_W}(x|\lambda) = G^2_s(x-1) p_{M_\mathcal{E}}(y=1|\lambda) + G^2_s(x) p_{M_\mathcal{E}}(y=0|\lambda), \tag{9}$$

where $p_{M_\mathcal{E}}(y|\lambda)$ is the response function of the measurement $[y|M_\mathcal{E}]$. Similarly, the operational equivalence of Eq. (7) implies that noncontextual models satisfy

$$p_\mathcal{M}(\lambda'|\lambda) = (1-p_d) p_\mathcal{I}(\lambda'|\lambda) + p_d p_{\mathcal{M}^D}(\lambda'|\lambda), \tag{10}$$

where $p_\mathcal{I}(\lambda'|\lambda)$ and $p_{\mathcal{M}^D}(\lambda'|\lambda)$ are matrices of transition probabilities representing the channels $\mathcal{I}$ and $\mathcal{M}_\mathcal{D}$ in the ontological model.

D. AWVs and contextuality beyond idealizations

In this section we will start our investigation by presenting two results. First, the assumption of noncontextuality limits the maximum value achievable by the quantity

$$p_- := \int_{-\infty}^{0} p(x, y=1|P_*, \mathcal{M}^W, M_F) dx, \tag{11}$$

even beyond the idealized scenario studied in Ref. [9]. Second, in the quantum treatment the relation between p_- and the weak value presented in Eq. (8) extends to situations where noise and imperfections are present. Combining these two results we obtain our first proof that (real) anomalous weak values are nonclassical beyond the idealized setting of Ref. [9]. What is more, we can quantify how strong the anomaly needs to be, for given noise, to prove contextuality.

To highlight the independence of our noncontextuality theorems from the quantum formalism, we introduce the notation $\simeq$ to denote operationally indistinguishable procedures, following Ref. [10]. For example, instead of the operator equality of Eq. (5) we will write

$$[x|M_W] \simeq G^2_s(x-1)[y=1|M_\mathcal{E}] + G^2_s(x)[y=0|M_\mathcal{E}],$$

which means that the above two measurement procedures give rise to the same operational statistics for every preparation procedure taken as input. Similarly, Eq. (7) becomes

$$\mathcal{M} \simeq (1-p_d)\mathcal{I} + p_d \mathcal{M}^D,$$

denoting that, for any preparation procedure used to initialize the system, if we apply either of the above two transformations and then measure according to an arbitrary measurement procedure, the outcome statistics will be identical. When the relevant operational data arise from quantum experiments, however, $\simeq$ can be simply identified with the corresponding operator identities, as we did in the previous section.

Theorem 1 (noise-robust contextuality from the real part of the weak value). Suppose we have a noncontextual ontological model and that

(1) there exists a 2-outcome measurement $M_\mathcal{E}$ and a probability distribution $q(x)$ with median $x = 0$ such that, for all $x \in \mathbb{R}$,

$$[x|M_W] \simeq q(x-1)[y=1|M_\mathcal{E}] + q(x)[y=0|M_\mathcal{E}]; \tag{12}$$

(2) if $\mathcal{M} := \int \mathcal{M}^W_x dx$, there exists $p_d \in [0, 1]$ such that

$$\mathcal{M} \simeq (1-p_d)\mathcal{I} + p_d \mathcal{M}^D, \tag{13}$$

where $\mathcal{I}$ denotes the identity transformation and $\mathcal{M}^D$ some other transformation.

Then, if $p_- := \int_{-\infty}^0 p(x, y=1|P_*, \mathcal{M}^W, M_F) dx$ and $p_F := p(y=1|P_*, M_F)$,

$$p_- \leqslant p^{NC}_- := p_F \tfrac{1}{2} + (1-p_F) p_d. \tag{14}$$

Remarks. It follows from the first assumption that the marginal probability of the weak measurement M_W giving a negative result is at most $\frac{1}{2}$. If the system was totally undisturbed then the postselection would occur independently with probability p_F. This would give a joint probability of negative result and postselection of at most $\frac{p_F}{2}$. Our inequality shows that noncontextual models cannot explain measurement disturbance increasing the joint probability p_- above this no-disturbance bound by more than $O(p_d)$.

We provide the proof of this theorem in Appendix B. However, to give some intuition we give here a simplified proof that holds for a finite ontic state space and only derives a weaker noncontextuality bound (but still strong enough that it suffices to prove that real anomalous weak values are contextual).

Proof (simplified version). In the ontological model

$$p_- = \int_{-\infty}^{0} \sum_{\lambda',\lambda} p_{M_F}(y=1|\lambda') p_{\mathcal{M}^W}(x,\lambda'|\lambda) p_{P_*}(\lambda) dx \quad (15)$$

and

$$p_F = \sum_{\lambda} p_{M_F}(y=1|\lambda) p_{P*}(\lambda).$$

As discussed in Sec. II C, $p_{M_W}(x|\lambda) = \sum_{\lambda'} p_{\mathcal{M}^W}(x,\lambda'|\lambda)$. Hence $p_{M_W}(x|\lambda) \geqslant p_{\mathcal{M}^W}(x,\lambda|\lambda)$. Using measurement noncontextuality and Eq. (12), one obtains Eq. (9), i.e.,

$$p_{M_W}(x|\lambda) = G_s^2(x-1) p_{M_\mathcal{E}}(y=1|\lambda) + G_s^2(x) p_{M_\mathcal{E}}(y=0|\lambda).$$

Since $G_s^2(x)$ has median zero, this immediately implies

$$\int_{-\infty}^{0} p_{M_W}(x|\lambda) dx \leqslant \frac{p_{M_\mathcal{E}}(y=1|\lambda) + p_{M_\mathcal{E}}(y=0|\lambda)}{2} = \frac{1}{2}.$$

Hence, for the terms in Eq. (15) with $\lambda' = \lambda$ we have

$$\begin{aligned}
&\int_{-\infty}^{0} \sum_{\lambda} p_{M_F}(y=1|\lambda) p_{\mathcal{M}^W}(x,\lambda|\lambda) p_{P_*}(\lambda) dx \\
&\quad \leqslant \int_{-\infty}^{0} \sum_{\lambda} p_{M_F}(y=1|\lambda) p_{M_W}(x|\lambda) p_{P_*}(\lambda) dx \\
&\quad \leqslant \frac{1}{2} \sum_{\lambda} p_{M_F}(y=1|\lambda) p_{P_*}(\lambda) = \frac{p_F}{2}.
\end{aligned}$$

Furthermore, as discussed in Sec. II C, $p_\mathcal{M}(\lambda'|\lambda) = \int_{-\infty}^{\infty} p_{\mathcal{M}^W}(x,\lambda'|\lambda)$, hence $p_\mathcal{M}(\lambda'|\lambda) \geqslant \int_{-\infty}^{0} p_{\mathcal{M}^W}(x,\lambda'|\lambda) dx$. By Eq. (13) and transformation noncontextuality we have Eq. (10), i.e.,

$$p_\mathcal{M}(\lambda'|\lambda) = (1-p_d) p_\mathcal{I}(\lambda'|\lambda) + p_d p_{\mathcal{M}^D}(\lambda'|\lambda).$$

Then, since $p_\mathcal{I}(\lambda'|\lambda) = \delta_{\lambda'\lambda}$ (e.g., using noncontextuality and taking into account that $\mathcal{I}$ can be implemented by letting no time pass, so that no dynamics can occur), one has that for $\lambda' \neq \lambda$, $p_\mathcal{M}(\lambda'|\lambda) = p_d p_{\mathcal{M}^D}(\lambda'|\lambda)$. Hence, for the terms of Eq. (15) with $\lambda' \neq \lambda$ we have

$$\begin{aligned}
&\int_{-\infty}^{0} \sum_{\lambda} \sum_{\lambda' \neq \lambda} p_{M_F}(y=1|\lambda') p_{\mathcal{M}^W}(x,\lambda'|\lambda) p_{P_*}(\lambda) dx \\
&\quad \leqslant \sum_{\lambda} \sum_{\lambda' \neq \lambda} p_{M_F}(y=1|\lambda') p_\mathcal{M}(\lambda'|\lambda) p_{P_*}(\lambda) \\
&\quad = p_d \sum_{\lambda} \sum_{\lambda' \neq \lambda} p_{M_F}(y=1|\lambda') p_{\mathcal{M}^D}(\lambda'|\lambda) p_{P_*}(\lambda) \\
&\quad \leqslant p_d \sum_{\lambda} \sum_{\lambda' \neq \lambda} p_{\mathcal{M}^D}(\lambda'|\lambda) p_{P_*}(\lambda) \\
&\quad \leqslant p_d \sum_{\lambda} p_{P_*}(\lambda) \\
&\quad = p_d.
\end{aligned}$$

Summing the $\lambda' = \lambda$ and $\lambda' \neq \lambda$ terms gives $p_- \leqslant p_F/2 + p_d$. ■

Our first illustration of how this theorem operates is in the idealized scenario discussed above. First, the operational equivalences in Eq. (12) and Eq. (13) are satisfied with $q(x) = G_s^2(x)$, due to Eq. (5) and Eq. (7), respectively. Furthermore, $p_d = \frac{1-e^{-1/4s^2}}{2} = o(1/s^2)$. Hence, from the above theorem, the data can only be explained by a noncontextual ontological model if the probability p_- of passing the postselection and displaying a negative pointer position is

$$p_- \leqslant p_F/2 + o(1/s). \quad (16)$$

However, quantum mechanically $p_- = p_-^{\text{ideal}}$ as given by Eq. (8). When $\text{Re}(\langle \Pi_\phi \mathcal{E} \rangle_{\rho_*}) \geqslant 0$, p_- is always smaller than $p_F/2$ for s large enough. However, whenever $\text{Re}(\langle \Pi_\phi \mathcal{E} \rangle_{\rho_*}) < 0$ (anomalous real weak value) there exists an s large enough for which $p_-^{\text{ideal}} > p_F/2 + o(1/s)$, from which we obtain a proof of contextuality.

Note already that this statement does not require the preparation to be pure, as is the case in standard formulations. However, going beyond this, our theorem does not require the postselection to be exactly projective either. For example, let us assume that unbiased noise is present in the postselection, i.e., in the quantum description,

$$\begin{aligned}
\{[y=1|M_F], [y=0|M_F]\} &= (1-2\epsilon)\{\Pi_\phi, \mathbb{1} - \Pi_\phi\} \\
&\quad + 2\epsilon\{\mathbb{1}/2, \mathbb{1}/2\}, \quad (17)
\end{aligned}$$

where $\epsilon \in (0, \frac{1}{2})$. We show in Appendix D that the operational equivalences of Eqs. (12) and (13) are still satisfied and, furthermore,

$$p_- = p_-^{\text{noisy}} := \frac{p_F}{2} - \frac{1}{\sqrt{\pi} s} \text{Re}(\langle [y=1|M_F] \mathcal{E} \rangle_{\rho_*}) + o\left(\frac{1}{s}\right). \quad (18)$$

Hence, if $p_-^{\text{noisy}} > p_-^{NC} = p_F/2 + o(1/s)$ the experiment still provides a proof of contextuality. As is intuitive, p_-^{noisy} is determined by a *noisy weak value*, whose relation with the ideal one can be inferred from

$$\text{Re}(\langle [y=1|M_F] \mathcal{E} \rangle_{\rho_*}) = (1-2\epsilon) \text{Re}(\langle \Pi_\phi \mathcal{E} \rangle_{\rho_*}) + \epsilon p_\mathcal{E},$$

where $p_\mathcal{E} := \text{Tr}(\mathcal{E} \rho_*)$. This clarifies that the noise, parametrized by ϵ, linearly "damps" the potential negativity of the weak value. In fact, using $\text{Re}(\langle \Pi_\phi \mathcal{E} \rangle_{\rho_*}) \geqslant -1/8$ (Eq. (41) of Ref. [22]), we can estimate the noise threshold for $p_-^{\text{noisy}} > p_-^{\text{NC}}$ in Theorem 1 to be $\epsilon < \frac{1}{2+8p_\mathcal{E}}$.

As an experimental proposal, one can consider the setup of Ref. [11]. The measured p_-^{noisy} is well above p_-^{NC}. Hence, if Eqs. (12) and (13) were verified [only Eq. (12) is claimed], the experiment would be a proof of contextuality from AWVs. The importance of checking all the operational equivalences $\simeq$ will be stressed later (Sec. V), when we show that, if even one of them is dropped, a classical model exists reproducing the anomaly.

We conclude this section by discussing in more detail the relation between Theorem 1 and the main theorem of Ref. [9]. One can note that Eq. (12) is exactly the first operational equivalence used in Ref. [9], while Eq. (13) is a stronger operational requirement than the second equivalence of Ref. [9], as Eq. (13) involves the transformation rather than the measurement. Importantly, Theorem 1 makes no reference to the properties of $[y|M_F]$ (for example, we do not require any of the properties associated with projective measurements

in quantum theory), which is what allowed the above discussion of experimental proofs of contextuality from AWVs in nonideal scenarios. As a minor difference, the inequality derived in Ref. [9] is[3] $p_- \leqslant \frac{1}{2}p_F + p_d$, whereas we now obtain $p_- \leqslant \frac{1}{2}p_F + (1-p_F)p_d$. Since $0 < 1 - p_F < 1$ our bound is strictly stronger, although because p_d will typically be very small the improvement is minor. We will later provide evidence that the improved bound is tight.

E. Contextuality from imaginary weak value

Our second theorem concerns the imaginary part of the weak value. The theorem of Ref. [9] does not imply any connection between $\mathrm{Im}(_\phi\langle O\rangle_\rho) \neq 0$ and contextuality; furthermore, the imaginary weak value has analogs in classical models [12,13]. Nevertheless, we show that quantum-mechanical imaginary weak values are contextual. This complements the results of Ref. [9] by showing that *every* anomalous weak value is nonclassical—not just those with an anomalous real part.

Let us recall how the imaginary part of a weak value is accessed experimentally. Suppose we keep the same initial pointer state $|\Psi\rangle_P$ as Eq. (2) and same interaction Hamiltonian $H = \mathcal{E} \otimes \Gamma$ applied for a unit time ($\mathcal{E}$ is some projector and Γ the momentum operator of the pointer). However, the pointer is measured in the momentum basis $\{|\gamma\rangle\}$. This gives a POVM $[\gamma|M_W] = N_\gamma^\dagger N_\gamma$ on the system with

$$N_\gamma = \langle\gamma|e^{-iH}|\Psi\rangle_P = \langle\gamma|\Psi\rangle_P\,[\exp(-i\gamma)\mathcal{E} + \mathcal{E}^\perp], \tag{19}$$

so that

$$[\gamma|M_W] = |\langle\gamma|\Psi\rangle_P|^2(\mathcal{E} + \mathcal{E}^\perp) = |\langle\gamma|\Psi\rangle_P|^2\mathbb{1}, \tag{20}$$

with $\langle\gamma|\Psi\rangle_P = \pi^{-1/4}\sqrt{s}\exp(-\frac{p^2s^2}{2})$. Note that these are exactly the POVM elements for a trivial measurement sampling from the probability distribution $|\langle\gamma|\Psi\rangle_P|^2$, which has median zero.

The choice of measurement on the pointer does not affect the marginal channel on the system and so Eq. (7) is still satisfied with $p_d = o(1/s)$:

$$\begin{aligned}\mathcal{M}(\cdot) &= \int_{-\infty}^{+\infty} d\gamma\,\mathcal{M}_\gamma^W(\cdot) = \int_{-\infty}^{+\infty} d\gamma\,N_\gamma(\cdot)N_\gamma^\dagger \\ &= (1-p_d)\mathcal{I}(\cdot) + p_d\mathcal{M}^D(\cdot).\end{aligned} \tag{21}$$

Furthermore, we show in Appendix E that in the ideal case we obtain a negative momentum and successful postselection with probability

$$p_-^{\text{ideal}} = \frac{p_F}{2} - \frac{1}{\sqrt{\pi}s}\,\mathrm{Im}(\langle\Pi_\phi\mathcal{E}\rangle_{\rho_*}) + o\left(\frac{1}{s}\right). \tag{22}$$

Given the above setting, that nonzero imaginary values of the weak values are a proof of contextuality is a consequence of the following theorem, proven in Appendix B:

Theorem 2. Suppose we have a noncontextual ontological model and that

(1) if M_{triv} involves ignoring the system and sampling a $\gamma \in \mathbb{R}$ that is negative with probability $\frac{1}{2}$,

$$[\gamma|M_W] \simeq [\gamma|M_{\text{triv}}]; \tag{23}$$

(2) if $\mathcal{M} := \int \mathcal{M}_\gamma^W d\gamma$, there exists $p_d \in [0,1]$ such that

$$\mathcal{M} \simeq (1-p_d)\mathcal{I} + p_d\mathcal{M}^D, \tag{24}$$

where $\mathcal{I}$ denotes the identity transformation and $\mathcal{M}^D$ some other transformation.

Then if $p_- := \int_{-\infty}^0 p(\gamma, y=1|P_*, \mathcal{M}^W, M_F)d\gamma$ and $p_F := p(y=1|P_*, M_F)$,

$$p_- \leqslant p_-^{\text{NC}} = p_F\tfrac{1}{2} + (1-p_F)p_d. \tag{25}$$

Remarks. Note that the result requires no mention of $M_\mathcal{E}$. The assumptions of the theorem are satisfied by the experimental setting measuring the imaginary part of the weak value. In fact, Eq. (23) and Eq. (24) follow immediately from Eq. (20) and Eq. (21), respectively. Hence, since $p_d = o(1/s)$, as before if we observe $p_- > p_-^{\text{NC}} = p_F\frac{1}{2} + o(1/s)$ we have a proof of contextuality. From Eq. (22) this happens whenever $\mathrm{Im}(_\phi\langle\mathcal{E}\rangle_\rho)$ is negative (recall from Sec. II B that a nonzero imaginary part can be taken negative without loss of generality). Hence, imaginary weak values are contextual. Together with the theorem of the previous section, this shows that *all* (real or imaginary) anomalous weak values are contextual. The theorem also covers noisy postselection in exactly the same way as we discuss after Theorem 1, with p_-^{ideal} of Eq. (22) substituted by a noisy analog involving a "noisy imaginary weak value," as discussed in Eq. (18) for the real part.

The status of imaginary weak values

This result contrasts with the dismissal of the imaginary parts of weak values in Ref. [9]. The discussion there begins by pointing out that "the imaginary part [...] is manifested very differently from the real part [15]."

This is true, and explains why the proof of contextuality has to be adapted slightly to apply to this case. More formally, we could note that the relevant Kraus operators of the weak measurement on the system when we access the pointer's momentum are proportional to unitaries $\exp(-i\mathcal{E}\gamma)$ [see Eq. (19)]. Hence, the same instrument could be achieved by classically sampling an "outcome" γ (as in M_{triv} above) and then directly performing the appropriate unitary. When we do things this way, it is clear that the correlation between the sampled outcome and the postselection is purely due to the fact we have disturbed the system by applying a unitary. Yet, since the same instrument is implemented as in the measurement of the imaginary part of the weak value, the same proof of contextuality holds for this sampling scheme. Nonclassicality arises in this case because the unitaries are strong enough to significantly affect the postselection and yet they average out to something very close to the identity channel. While the leading-order effect of the unitaries on the postselection is captured by exactly the imaginary part of the weak value, if one has actually implemented the instrument by applying various unitaries it is unclear why this

[3]Notice that what was called p_- in [9] is what we call $\frac{p_-}{p_F}$ here.

should be expected to reveal anything about the "value" of the system observable.

This brings us to the next sentence of Ref. [9], which gives a specific argument against imaginary weak values being nonclassical: "Indeed complex weak values are easily obtained even in the Gaussian subset of quantum mechanics, which has weak measurements (with the same information-tradeoff disturbance [*sic*] utilized here) and yet admits a very natural noncontextual model [12]."

Weak measurements in the referenced model have since been explored in detail by Karanjai *et al.* [13]. The model gives definite values to all the allowed observables and so one can meaningfully talk about what values observables truly have independently of any measurement. It is found in Ref. [13] that the real part of the weak value reflects the true average value of the observable given the information from the preparation and postselection. Imaginary parts can also arise, but they are purely an artifact of disturbance, in agreement with the discussion above.

Since the model in Ref. [12] is noncontextual, the Gaussian subset of quantum mechanics cannot violate any noncontextuality inequalities. But the weak values in the theory do have imaginary parts, and the weak measurements thereof satisfy Eq. (23). Therefore the measurements must fail to satisfy Eq. (24) with a sufficiently small p_d. In other words, if we measure disturbance using p_d then, contrary to the claim in parentheses in the quotation above, the weak measurements considered in Ref. [13] do *not* have the favorable information-disturbance tradeoff needed to prove contextuality.

We should clarify that this is not in contradiction with our calculations of p_d because those calculations are only valid for the weak measurement of a projector, which has eigenvalues 0 and 1. The only observables that can be weakly measured in Ref. [12] are linear combinations of position and momentum operators, which all have unbounded spectrum. To get some intuition for why this makes a difference, we can easily generalize the calculation of $\mathcal{M}$ to the case of measuring an operator $O = \sum_i o_i \mathcal{E}_i$ with an arbitrary finite number of eigenvalues $\{o_i\}$, giving

$$\mathcal{M}(\rho) = \sum_{i,j} \exp\left(-\frac{(o_i - o_j)^2}{4s^2}\right) \mathcal{E}_i \rho \mathcal{E}_j. \tag{26}$$

Since $\mathcal{I}(\rho) = \sum_{i,j} \mathcal{E}_i \rho \mathcal{E}_j$, to satisfy Eq. (24) we must have $\mathcal{M}^D(\rho) = \sum_{i,j} C_{ij} \mathcal{E}_i \rho \mathcal{E}_j$ with

$$C_{ij} = \frac{1}{p_d}\left[\exp\left(-\frac{(o_i - o_j)^2}{4s^2}\right) - (1 - p_d)\right]. \tag{27}$$

Notice that the Choi-Jamiolkowski state associated with $\mathcal{M}^D$ has a block-diagonal structure in which C_{ij} appear. Hence, $\mathcal{M}^D$ is completely positive if and only if C_{ij} are the entries of a positive matrix. In particular this requires $|C_{ij}| \leqslant \frac{C_{ii}+C_{jj}}{2} = 1$, where $C_{ii} = 1$ follows directly from Eq. (27). The requirements that $C_{ij} \geqslant -1$ for all (i, j) can be written

$$p_d \geqslant \frac{1}{2}\left[1 - \min_{i,j} \exp\left(-\frac{(o_i - o_j)^2}{4s^2}\right)\right]. \tag{28}$$

Hence as we increase the difference between the smallest and largest o_i, we need a larger s to ensure a small p_d. This suggests that for operators with an unbounded spectrum we should expect that Eq. (24) can only be satisfied with $p_d \geqslant \frac{1}{2}$, which is far too large to allow a violation of the noncontextuality inequality in Eq. (25).

F. AWVs and contextuality with qubit pointers or coarse graining

While Theorem 1 removed the idealizations of a perfectly projective postselection and pure input states from the main result of Ref. [9], we still followed the traditional approach of introducing weak values using a continuous-variable pointer; see Sec. II B. Correspondingly, Theorem 1 strictly requires an infinite number of operational equivalences to be satisfied, which cannot be checked by finite means. In the following, we will solve this issue.

It is known that one can follow an experimental setting for measuring weak values that is analogous to the one discussed above but uses a qubit pointer only [23]; alternatively, one can consider a coarse graining of x in the traditional setting of Sec. II B. Either way, in these alternative scenarios with finite degrees of freedom we are able to prove that (1) the connection between AWVs and contextuality holds and (2) as opposed to Theorem 1, the no-go theorem only requires us to verify a finite number of operational equivalences. The relevant no-go theorem, proven in Appendix B, is given by the following:

Theorem 3 (noise-robust no-go theorem: finite version). Suppose we have a noncontextual ontological model and that

(1) there exists a measurement $M_\mathcal{E}$ and a probability p_m such that

$$[x|M_W] \simeq p_m[x|M_\mathcal{E}] + (1 - p_m)[x|M_{\text{triv}}], \tag{29}$$

where M_{triv} involves ignoring the system and sampling an x that is negative with probability $\frac{1}{2}$;

(2) if $\mathcal{M} := \int \mathcal{M}^W_x dx$, there exists $p_d \in [0, 1]$ such that

$$\mathcal{M} \simeq (1 - p_d)\mathcal{I} + p_d \mathcal{M}^D, \tag{30}$$

where $\mathcal{I}$ denotes the identity transformation and $\mathcal{M}^D$ some other transformation.

Then if $p_- := \int_{-\infty}^{0} p(x, y = 1|P_*, \mathcal{M}^W, M_F)dx$ and $p_F := p(y = 1|P_*, M_F)$,

$$p_- \leqslant p_F \frac{1 + p_m}{2} + (1 - p_F)p_d. \tag{31}$$

Remarks. In Appendix F we describe a weak measurement scheme using a qubit pointer with small parameter ϵ. The outcome is a discrete $x = \pm 1$ so the integrals over x above reduce to sums. We show that the operational equivalences of Eq. (29) and Eq. (30) are satisfied with $p_m = 2\epsilon + o(\epsilon)$ and $p_d = o(\epsilon)$, respectively, and calculate

$$p_- = p_F \frac{1 + p_m}{2} - 2\epsilon \,\text{Re}(\langle \Pi_\phi \mathcal{E} \rangle_\rho) + o(\epsilon), \tag{32}$$

giving contextuality for sufficiently small ϵ whenever $\text{Re}(\langle \Pi_\phi \mathcal{E} \rangle_\rho) < 0$, as before.

The same argument can be made for the standard quantum experiment described in Sec. II B, once we coarse-grain the

pointer position to a two-outcome measurement M_W^{coarse} with outcomes $x \leqslant 1/2$ and $x \geqslant 1/2$ (i.e., x closest to the eigenvalue 0 of $\mathcal{E}$, or closest to the eigenvalue 1). If we now label these outcomes $x = -1$ and $x = +1$, respectively, then the conditions of Theorem 3 are satisfied with $p_d = o(1/s)$ and $p_m = 1/(\sqrt{\pi}s) + o(1/s)$. Then, for the perfect postselection,

$$p_- = p_F \frac{1+p_m}{2} - \frac{1}{\sqrt{\pi}s}\,\mathrm{Re}(\langle \Pi_\phi \mathcal{E}\rangle_{\rho_*}) + o\left(\frac{1}{s}\right), \tag{33}$$

which, with large s, violates the noncontextuality bound.

G. A remark on the debate concerning AWVs

Theorem 3 not only tells us that weak-value experiments proving contextuality can be conducted with qubit pointers, but also clarifies another issue of the weak-value debate. When Ferrie and Combes presented *discrete* classical toy models reproducing certain aspects of AWVs [5], questions were posed as to whether these are a good analog of the weak value due the intrinsic discreteness [6] (as opposed to the standard quantum experiment which is continuous or, when discrete, it is a coarse graining of a continuous measurement [24]). Theorem 3 shows that the contextuality of the weak value has nothing to do with the fact that we are performing a measurement of a continuous quantity—the pointer position or momentum: nonclassicality is present both in the coarse graining of the standard experiment as well as in an intrinsically discrete experiment. In particular, although the weak value no longer appears simply as an average pointer position, the correct "scaling procedure" to determine whether a discrete outcome is sufficiently biased to be considered anomalous can be determined operationally using p_m.

III. AN ALTERNATIVE APPROACH TO THE NO-GO THEOREMS

A. Theorem based on measurement and preparation noncontextuality

In Theorems 1–3 we removed the idealization of exact postselection in Ref. [9] and extended an operational equivalence on a measurement to a correspondent operational equivalence on a transformation, Eq. (13). In fact, Eq. (13) requires us to check that *every* subsequent measurement on the system is affected little by the weak measurement, whereas the original assumption only required checking that the postselection is affected little when preceded by the weak measurement. Here we present an alternative approach in which we keep the original, less demanding assumptions of Ref. [9] but we introduce some extra preparations whose aim is to provide an operational measure of how "close to projective" the postselection is.[4]

To do so, we do the following:

(1) We introduce an ensemble of preparations $[b|S]$, where $[b = 0|S]$ is prepared with probability q_0 and $[b = 1|S]$ is prepared with probability $q_1 = 1 - q_0$. In practice, we will look for S that maximizes the correlations with the corresponding outcomes of the (imperfect) postselection, i.e., maximizing

$$C_S := p(b = 0, y = 0|S, M_F) + p(b = 1, y = 1|S, M_F),$$

where $p(b, y|S, M_F)$ is the probability that $[b|S]$ is prepared and an immediate measurement of $[y|M_F]$ on $[b|S]$ returns outcome y.

(2) If P_* denotes the input preparation in the standard setting (as in Sec. II C), we include it in an ensemble where P_* is prepared with probability q_* and $P_\perp$ is prepared with probability $q_\perp = 1 - q_*$. $P_\perp$ and q_* are chosen such that $q_0[b = 0|S] + q_1[b = 1|S] \simeq q_* P_* + q_\perp P_\perp$.

It is useful to spell out what this means in quantum terms when the system being weakly measured is a qubit. We start with $\{M_F, \mathbb{1} - M_F\}$, the imperfect postselection POVM, and the preparation ρ_*. Then we look for states σ_b, $b = 0, 1$, that maximize $\mathrm{Tr}(M_F\sigma_1)$ and $\mathrm{Tr}[(\mathbb{1} - M_F)\sigma_0]$. We then need to find suitable q_b, q_*, and $\rho_\perp$ such that $q_*\rho_* + q_\perp\rho_\perp = q_0\sigma_0 + q_1\sigma_1$, to satisfy the correspondent operational equivalence. Note that if we accessed perfect postselections and preparations, then we would get $C_S = 1$ by choosing $\sigma_1 = |\phi\rangle\langle\phi|$ and $\sigma_0 = \mathbb{1} - |\phi\rangle\langle\phi|$. In practice the postselection is not exactly projective and σ_b will never be exactly pure, so that $C_S < 1$ experimentally.

We are now able to formulate a no-go theorem using this second strategy. Denoting by $p(x, y|P_*, M_F \circ M_W)$ the probability that if the system is initialized through the preparation procedure P_* and $[x|M_W]$, $[y|M_F]$ are sequentially measured, one obtains outcomes (x, y), we have the following:

Theorem 4 (noncontextuality inequality based on preparation noncontextuality). Suppose we have a noncontextual ontological model and that

(1) there exists a 2-outcome measurement $M_\mathcal{E}$ and a probability distribution $q(x)$ with median $x = 0$ such that, for all $x \in \mathbb{R}$,

$$[x|M_W] \simeq q(x-1)[y = 1|M_\mathcal{E}] + q(x)[y = 0|M_\mathcal{E}]; \tag{34}$$

(2) given the sequential measurement $[x, y|M_F \circ M_W]$, we define $[y|\tilde{M}_F] := \int dx [x, y|M_F \circ M_W]$; then there exists $p_d \in [0, 1]$ such that

$$[y|\tilde{M}_F] \simeq (1 - p_d)[y|M_F] + p_d[y|M_D], \tag{35}$$

for some 2-outcome measurement $[y|M_D]$;

(3) there exists an ensemble

$$\{\{q_*, P_*\}, \{q_\perp, P_\perp\}\},$$

such that

$$q_0[b = 0|S] + q_1[b = 1|S] \simeq q_* P_* + q_\perp P_\perp. \tag{36}$$

Then, if $p_- := \int_{-\infty}^{0} p(x, y = 1|P_*, M_F \circ M_W)dx$ and $p_F := p(y = 1|P_*, M_F)$,

$$p_- \leqslant p_F \frac{1}{2} + (1 - p_F)p_d + \frac{1 - C_S}{2q_*}. \tag{37}$$

Remarks. The theorem is proved in Appendix C. It parallels Theorem 1; in particular Eq. (34) is the same as Eq. (12). Because the logical structure of Appendix C parallels that of Appendix B, theorems parallel to Theorems 2 and 3 can

[4]This general strategy to "robustify" contextuality proofs was first proposed in Ref. [25].

also be proven, with the assumption in Eq. (24) and Eq. (30) replaced by the conjunction of Eq. (35) and Eq. (36).

Now that the alternative theorem is stated, let us discuss in more detail the differences with our first approach by comparing Theorem 4 with Theorem 1. The requirement Eq. (34) is exactly the same; the operational equivalence of Eq. (35) is strictly weaker than the correspondent Eq. (13), since the latter requires us to verify that the weak measurement affects only slightly *any* subsequent measurement, whereas the former only requires us to check the same condition for the postselection measurement M_F; the operational equivalence of Eq. (36) is added, and involves the addition of preparations S used for testing the quality of the postselection, as well as of a preparation $P_\perp$ that provides a nontrivial operational equivalence; finally, the bound on p_- matches the analog one from Theorem 1, with an extra punishing term proportional to $1 - C_S$. The bound hence becomes increasingly weak as the postselection departs from the perfect predictability associated with projective measurements in quantum theory.

B. Application: Assessing current AWV experiments

The second version of the theorem can also be compared with the quantum-mechanical predictions. For example, in the unbiased noise model presented in the previous section one can show that all the operational equivalences of Theorem 4 are satisfied. Furthermore, one can use Eq. (18) and note that $C_S = 1 - \epsilon$ (see Appendix D).

We can once more compare with the experimental setting of Ref. [11]. First, note that only the operational equivalences of Eqs. (34) and (35) are claimed, so one would need to complete this with Eq. (36) to get that the violation of the bound of Eq. (37) is a proof of contextuality. In other words, in principle the same data can be utilized by simply adding an estimation of the sharpness of the postselection through an extra preparation satisfying Eq. (36). We can, in fact, work out from the experimental data how close to projective the postselection needs to be for the claim of contextuality from AWVs of Ref. [11] to hold. Specifically, one has $p_d = 0.0019$, $s = 8.10336$, $p_F = 0.0475865$, $\frac{p_-}{p_F} = 0.602927$, and (with the obvious choice of fair ensembles) $q_* = 1/2$. One can then estimate that Eq. (37) is satisfied only if $C_S > 0.996912$. We see that in this case the postselection needs to be very close to ideal.[5]

IV. REMAINING IDEALIZATIONS: PERFECT OPERATIONAL EQUIVALENCES

Some readers may have noticed that there is an idealization that was not dealt with in Theorems 1–4. That is, any experiment will only ever verify the operational equivalences $\simeq$ up to some approximation. Luckily, as discussed in Ref. [26], this can be dealt with using a generic technique. One begins by assuming access to a tomographically complete set of procedures that enables the operational equivalences to be checked. The basic idea is then that while the "primary" procedures (i.e., the ones actually implemented) will not satisfy the operational equivalence exactly, we can use their statistics to find "secondary" preparations in their convex hull[6] that do satisfy the equivalences exactly. It is to these secondary preparations that we can apply Theorems 1–4. In particular, as we discussed one can apply Theorem 3 both to the single qubit pointer experiment, as well as the coarse-grained version of the standard experiment—meaning that we only need to apply the above discussion to a finite set of operational equivalences. The price for using this technique is that the secondary procedures are more mixed than the primary ones and hence will give smaller values for C_S and p_-. In that sense, applying this technique builds upon the noise robustness to nonideal values of such parameters that we have provided here.

A last comment. The last remaining idealization at this point is that we assumed we know a tomographically complete set of measurements. Strictly, we cannot prove that a given set of measurement procedures is complete without relying on the quantum formalism. However, one can gather evidence from the experimental data that a given set is complete. This goes beyond the scope of the present work, but is discussed in detail in Ref. [26], and techniques to address this issue have since been introduced in Ref. [27].

V. NECESSITY OF OPERATIONAL EQUIVALENCES FOR NONCLASSICALITY OF AWVs

We have seen that the statistics collected by the AWV experiment cannot be reproduced by a noncontextual ontological model in the presence of some extra operational constraints:

(1) Eqs. (12) and (13) in the case of Theorem 1, with similar constraints for Theorems 2 and 3;

(2) Eqs. (34)–(36) in the case of Theorem 4.

At first sight this might seem rather involved, especially if compared to broader claims of nonclassicality of the AWVs that appeared in the literature. Here, however, we show that dropping *any* of the operational equivalences in any of the theorems allows for the explicit construction of a classical (noncontextual) ontological model that reproduces the anomaly. In fact, the models even reproduce the full quantum statistics of the sequential measurement on ρ_* and not just the anomaly of the pointer. Hence our conditions are not only sufficient, but they are also necessary, showing that AWVs can only be understood as unavoidably quantum in the presence of all the operational constraints described. Hopefully this will help in clarifying the debate that arose around this topic, showing that both "sides" are indeed correct: in a way similar to that in which nonlocal correlations are a quantum phenomenon only in a setting in which signaling has been excluded, AWVs are indeed fundamentally quantum, but only when accompanied by certain extra operational facts.

[5]Not unrealistically close. A quantity comparable to C_S was reported as 0.99709(7) in another contextuality experiment [26].

[6]In general "supplementary procedures" have to be implemented to ensure that the convex hull extends in all directions [26].

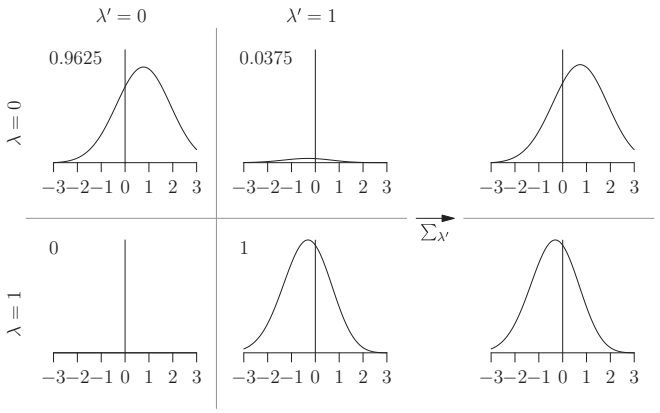

FIG. 3. An illustration of the model in Sec. V A 1. On the left are plots of $p_{\mathcal{M}^W}(x, \lambda'|\lambda)$ against x, and the numbers $p_{\mathcal{M}}(\lambda'|\lambda) = \int_{-\infty}^{\infty} p_{\mathcal{M}^W}(x, \lambda'|\lambda)dx$. On the right are plots of $p_{M_W}(x|\lambda) = \sum_{\lambda'} p_{\mathcal{M}^W}(x, \lambda'|\lambda)$ against x. The operational probabilities used are quantum probabilities from the standard scheme with parameters chosen so that $p_F = \frac{1}{5}$, $p_d = \frac{1}{20}$, and $_\phi\langle\mathcal{E}\rangle_\psi = -\frac{1}{2}$. (In particular those parameters include a rather small $s \approx 1.5$ to ensure that all features are visible. This s is still large enough for our noncontextuality inequalities to be violated.) Notice on the left that $\lambda = \lambda'$ with high probability, but on the right we see that the $\lambda = 1$ ontic state is predisposed to give negative values of x.

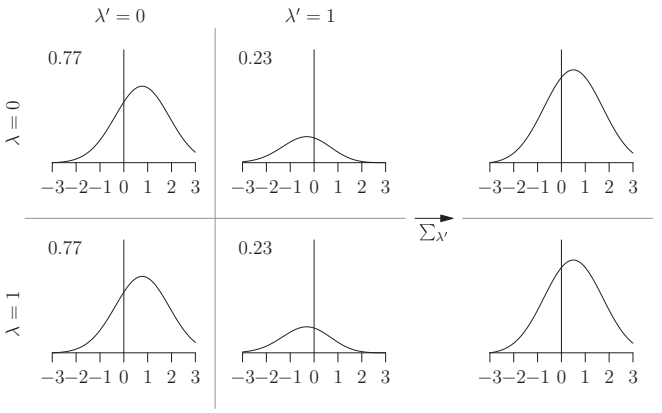

FIG. 4. As in Fig. 3, but for the model of Sec. V A 2. Notice on the right that neither ontic state is predisposed to give negative x, but on the left we see that the $\lambda = 1$ state is very likely to be disturbed to $\lambda' = 0$.

A. Necessity of conditions in Theorems 1–3

In both of the following models, we take the ontic state λ to be y, i.e., a determination of the outcome of M_F, and we set

$$p_{P_*}(\lambda) = p(y = \lambda|P_*, M_F). \tag{38}$$

1. Necessity of condition 1

The basic idea of our first model is to give results for the weak measurement according to the operational distribution under the predetermined postselection $y = \lambda$. That is, we set

$$p_{M_W}(x|\lambda) \approx p(x|P_*, \mathcal{M}^W, M_F, y = \lambda). \tag{39}$$

Exact equality in Eq. (39) would allow us to reproduce the operational distribution over x without any disturbance to the ontic state at all, at the price of violating the conditions on p_{M_W} arising from measurement noncontextuality (a failure of condition 1). However, we also want to reproduce the operational fact that whether or not the weak measurement is done affects the probabilities of M_F and so we add the minimal amount of disturbance necessary to achieve this. This amount of disturbance is, unsurprisingly, bounded by the p_d from Eq. (13). We then actually sample x from the operational distribution for $y = \lambda'$, the disturbed ontic state, which is why Eq. (39) is only approximately true. The model is illustrated in Fig. 3; for the full detail of how to implement the minimal disturbance see Appendix G.

2. Necessity of condition 2

This time we ignore λ and simply distribute (x, λ') according to the operational probabilities for (x, y), at the expense of a very large disturbance to the postselection:

$$p_{\mathcal{M}^W}(x, \lambda'|\lambda) = p(x, y = \lambda'|P_*, \mathcal{M}^W, M_F). \tag{40}$$

By construction

$$\sum_{\lambda'} p_{\mathcal{M}^W}(x, \lambda'|\lambda) = p(x|P_*, M_W), \tag{41}$$

so we satisfy any operational equivalences for M_W (condition 1 is satisfied).

Intuitively, notice that $\lambda = 1$ is greatly disturbed by the model since the probability of going to $\lambda' = 0$ is $p(y = 0|P_*, \mathcal{M}, M_F) \approx 1 - p_F$ (the probability of not passing the postselection). This is a failure of condition 2 whenever that probability exceeds p_d. These features can be seen in Fig. 4.

B. Necessity of conditions in Theorem 4

1. Necessity of condition 1

This follows from the first model above. To satisfy condition 3, we can set $p(\lambda|P) = p(y|P, M_F)$ for any preparation procedure P. This respects convexity and if two procedures are operationally equivalent they will in particular have the same $p(y|P, M_F)$ and hence the same $p(\lambda|P)$, as required by preparation noncontextuality.

2. Necessity of condition 2

This follows similarly from the second model above.

3. Necessity of condition 3

The final ontological model we consider is the ψ-complete model [28], which is well known to be measurement noncontextual. In fact we will consider the generalization of the ψ-complete model to an arbitrary operational theory. The set of ontic states λ is identified with the set of (convexly extremal) preparations, $p_P(\lambda) = \delta(\lambda - P)$, and the response functions are given by the operational probabilities, $p_M(x|\lambda) = p(x|P = \lambda, M)$. This model reproduces the operational probabilities and is measurement noncontextual (that is, satisfies conditions 1 and 2 of Theorem 4); however it does not satisfy preparation noncontextuality, since it does not associate the same distributions to the ensembles associated with S and $\{P_*, P_\perp\}$. Hence, condition 3 cannot be dropped from Theorem 4.

VI. CONCLUSIONS

Our results show that contextuality captures what is nonclassical about anomalous weak values in a way that is experimentally relevant and wide ranging. In particular, the postselection need not be a perfect projective measurement, the pointer need not be a continuous-variable system, and if there is an imaginary part to the weak value then the real part need not be anomalous.

On the other hand, we have shown through explicit noncontextual models that if any of the operational equivalences we use are absent a classical explanation is possible.

Our results also answer some of the questions left open in Ref. [21]. There, it was shown that the fluctuation theorem experiments probing the Margenau-Hills work quasiprobability introduced in Ref. [22] can witness contextuality. However, it was left open how to make the argument robust to experimental imperfections. Here we gave the tools to do so.

ACKNOWLEDGMENTS

M.F.P. thanks Jonathan Barrett and Sina Salek for useful discussions, particularly regarding the imaginary parts of weak values, and disturbance in Ref. [13]. M.L. thanks Fabrizio Piacentini and Marco Genovese for useful discussions. R.K. thanks Elie Wolfe for discussion on the algorithmic aspects. Research at Perimeter Institute is supported by the Government of Canada through the Department of Innovation, Science, and Economic Development Canada and by the Province of Ontario through the Ministry of Research, Innovation, and Science. M.F.P. is supported by the Royal Commission for the Exhibition of 1851. M.L. acknowledges financial support from the European Union's Marie Skłodowska-Curie individual Fellowships No. H2020-MSCA-IF-2017 and No. GA794842, Spanish MINECO (Severo Ochoa SEV-2015-0522 and Project No. QIBEQI FIS2016-80773-P), Fundacio Cellex, and Generalitat de Catalunya (CERCA program and SGR 1381).

APPENDIX A: IDEAL, STANDARD QUANTUM SCENARIO

The channel induced by the weak measurement when the outcome is not recorded is $\mathcal{M}(\cdot)=\int_{-\infty}^{+\infty}\mathcal{M}_x^W(\cdot)=\int_{-\infty}^{+\infty}N_x(\cdot)N_x^\dagger$. Using the integral $\int_{-\infty}^{+\infty}G_s(x-a)G_s(x-b)dx=e^{-(a-b)^2/4s^2}$ and Eq. (4), one finds for every ρ

$$\begin{aligned}\mathcal{M}(\rho)&=\mathcal{E}\rho\mathcal{E}+\mathcal{E}^\perp\rho\mathcal{E}^\perp+e^{-1/4s^2}(\mathcal{E}\rho\mathcal{E}^\perp+\mathcal{E}^\perp\rho\mathcal{E})\\&=\frac{1}{2}\rho+\frac{1}{2}(\mathcal{E}-\mathcal{E}^\perp)\rho(\mathcal{E}-\mathcal{E}^\perp)\\&\quad+e^{-1/4s^2}\left[\frac{1}{2}\rho-\frac{1}{2}(\mathcal{E}-\mathcal{E}^\perp)\rho(\mathcal{E}-\mathcal{E}^\perp)\right]\\&=\frac{1+e^{-1/4s^2}}{2}\rho+\frac{1-e^{-1/4s^2}}{2}(\mathcal{E}-\mathcal{E}^\perp)\rho(\mathcal{E}-\mathcal{E}^\perp)\\&=(1-p_d)\rho+p_d(\mathcal{E}-\mathcal{E}^\perp)\rho(\mathcal{E}-\mathcal{E}^\perp),\end{aligned}$$

with $p_d=\frac{1-e^{-1/4s^2}}{2}$. Hence, $\mathcal{M}=p_d\mathcal{I}+(1-p_d)\mathcal{M}_D$, with $\mathcal{M}_D(\rho):=(\mathcal{E}-\mathcal{E}^\perp)\rho(\mathcal{E}-\mathcal{E}^\perp)$. It is then clear that the operational equivalences required by Theorem 1 are satisfied in the ideal case.

Finally, one can compute $p_-^{\text{ideal}}=\int_{-\infty}^0 dx\,\mathrm{Tr}(\Pi_\phi N_x\rho_* N_x^\dagger)=\frac{p_F}{2}-\frac{\mathrm{Re}(\langle\Pi_\phi\mathcal{E}\rangle_{\rho_*})}{\sqrt{\pi}s}+o(\frac{1}{s})$. This is a simple calculation; see, e.g., the proof of Lemma 1 in Ref. [21] (note, however, that we redefined p_-^{ideal} without the normalization by the postselection probability).

APPENDIX B: PROOF OF THEOREMS 1–3 AND REMARKS ON TIGHTNESS OF THE INEQUALITIES

All three theorems follow from the same basic argument; hence it is convenient to formulate all of them as corollaries of the following technical lemma:

Lemma 1 (noncontextuality inequality template 1). Suppose we have a noncontextual ontological model and that

(1) for any input λ, the probability of a negative outcome of $[x|M_W]$ is bounded by some value independent of the ontic state:

$$\int_{-\infty}^0 p_{M_W}(x|\lambda)dx\leqslant\tilde{p};\tag{B1}$$

(2) if $\mathcal{M}:=\int\mathcal{M}_x^W dx$, there exists $p_d\in[0,1]$ such that

$$\mathcal{M}\simeq(1-p_d)\mathcal{I}+p_d\mathcal{M}^D,\tag{B2}$$

where $\mathcal{I}$ denotes the identity transformation and $\mathcal{M}^D$ some other transformation.

Then, if $p_-:=\int_{-\infty}^0 p(x,y=1|P_*,\mathcal{M}^W,M_F)dx$ and $p_F:=p(y=1|P_*,M_F)$,

$$p_-\leqslant p_F\tilde{p}+(1-p_F)p_d=:p_-^{\text{NC}}.\tag{B3}$$

Proof. Define $\Lambda_1^\lambda=\{\lambda':p_{M_F}(y=1|\lambda')\leqslant p_{M_F}(y=1|\lambda)\}$ (i.e., λ' is undisturbed or uselessly disturbed, in terms of probability of passing the postselection) and $\Lambda_2^\lambda=\Lambda\setminus\Lambda_1^\lambda=\{\lambda':p_{M_F}(y=1|\lambda')>p_{M_F}(y=1|\lambda)\}$ (i.e., λ' usefully disturbed). In the ontological model,

$$\begin{aligned}p_-&=\int_{-\infty}^0 p(x,y=1|P_*,\mathcal{M}^W,M_F)dx\\&=\int_{-\infty}^0\iint p_{M_F}(y=1|\lambda')\\&\quad\times p_{\mathcal{M}^W}(x,\lambda'|\lambda)p_{P_*}(\lambda)d\lambda'd\lambda dx.\end{aligned}\tag{B4}$$

As described in Sec. II C, $p_{M_W}(x|\lambda)=\int_\Lambda p_{\mathcal{M}^W}(x,\lambda'|\lambda)d\lambda'$. Also, note that $\int p_{M_F}(y=1|\lambda)p_{P_*}(\lambda)d\lambda=p_F$. Hence, for the Λ_1^λ part of (B4), using Eq. (B1),

$$\begin{aligned}&\int_{-\infty}^0 dx\int_\Lambda d\lambda\int_{\Lambda_1^\lambda}d\lambda' p_{M_F}(y=1|\lambda')p_{\mathcal{M}^W}(x,\lambda'|\lambda)p_{P_*}(\lambda)\\&\leqslant\int_{-\infty}^0 dx\int_\Lambda d\lambda\int_{\Lambda_1^\lambda}d\lambda' p_{M_F}(y=1|\lambda)p_{\mathcal{M}^W}(x,\lambda'|\lambda)p_{P_*}(\lambda)\\&=\int_{-\infty}^0 dx\int_\Lambda d\lambda p_{M_F}(y=1|\lambda)p_{M_W}(x|\lambda)p_{P_*}(\lambda)-c\\&\leqslant\tilde{p}\int_\Lambda p_{M_F}(y=1|\lambda)p_{P_*}(\lambda)d\lambda-c=\tilde{p}p_F-c,\end{aligned}\tag{B5}$$

with a "correction" term which measures the contribution to (B4) lost due to useless disturbance:

$$c = \int_{-\infty}^{0} dx \int_{\Lambda} d\lambda \int_{\Lambda_2^\lambda} d\lambda' p_{M_F}(y=1|\lambda) p_{\mathcal{M}^W}(x,\lambda'|\lambda) p_{P_*}(\lambda). \tag{B6}$$

As described in Sec. II C, $p_{\mathcal{M}}(\lambda'|\lambda) = \int_{-\infty}^{\infty} p_{\mathcal{M}^W}(x,\lambda'|\lambda)dx$. Hence, $p_{\mathcal{M}}(\lambda'|\lambda) \geqslant \int_{-\infty}^{0} p_{\mathcal{M}^W}(x,\lambda'|\lambda)dx$. By Eq. (B2) and (transformation) noncontextuality we have $p_{\mathcal{M}}(\lambda'|\lambda) = (1-p_d)p_{\mathcal{I}}(\lambda'|\lambda) + p_d p_{\mathcal{M}^D}(\lambda'|\lambda)$. Since $\mathcal{I}$ can be implemented, for example, by letting no time pass so that no dynamical evolution is possible, transformation noncontextuality requires $p_{\mathcal{I}}(\lambda'|\lambda) = \delta(\lambda'-\lambda)$. Hence, $\int d\lambda \int_{\Lambda_2^\lambda} d\lambda' [p_{M_F}(y=1|\lambda') - p_{M_F}(y=1|\lambda)] p_{\mathcal{I}}(\lambda'|\lambda) p(\lambda) = 0$. It follows that for the part of (B4) with $\lambda' \in \Lambda_2^\lambda$ we have

$$\begin{aligned}
&\int_{-\infty}^{0} dx \int_{\Lambda} d\lambda \int_{\Lambda_2^\lambda} d\lambda' p_{M_F}(y=1|\lambda') p_{\mathcal{M}^W}(x,\lambda'|\lambda) p_{P_*}(\lambda) \\
&= \int_{-\infty}^{0} dx \int_{\Lambda} d\lambda \int_{\Lambda_2^\lambda} d\lambda' [p_{M_F}(y=1|\lambda') \\
&\quad - p_{M_F}(y=1|\lambda)] p_{\mathcal{M}^W}(x,\lambda'|\lambda) p_{P_*}(\lambda) + c \\
&\leqslant \int_{\Lambda} d\lambda \int_{\Lambda_2^\lambda} d\lambda' [p_{M_F}(y=1|\lambda') \\
&\quad - p_{M_F}(y=1|\lambda)] p_{\mathcal{M}}(\lambda'|\lambda) p_{P_*}(\lambda) + c \\
&= p_d \int_{\Lambda} d\lambda \int_{\Lambda_2^\lambda} d\lambda' [p_{M_F}(y=1|\lambda') \\
&\quad - p_{M_F}(y=1|\lambda)] p_{\mathcal{M}^D}(\lambda'|\lambda) p_{P_*}(\lambda) + c \\
&\leqslant p_d \int_{\Lambda} d\lambda \int_{\Lambda_2^\lambda} d\lambda' [1 - p_{M_F}(y=1|\lambda)] p_{\mathcal{M}^D}(\lambda'|\lambda) p_{P_*}(\lambda) + c \\
&\leqslant p_d \int_{\Lambda} [1 - p_{M_F}(y=1|\lambda)] p_{P_*}(\lambda) d\lambda + c \\
&= (1-p_F)p_d + c.
\end{aligned}$$

Summing the Λ_1^λ and Λ_2^λ contributions gives $p_- \leqslant p_F \tilde{p} + (1-p_F)p_d$. ■

Our inequality is slightly tighter than the $p_- \leqslant p_F\tilde{p} + p_d$ one would expect from [9]. In order to check whether our inequality is in fact maximally tight, we applied the algorithmic approach to noncontextuality inequalities described in Ref. [14]. Since that approach requires fixed operational equivalences, we repeated this procedure for many numerical values of the parameters $\tilde{p}$, p_d and verified that our inequalities define facets of the corresponding "noncontextuality polytope" [14] in each case (see Appendix H). It appears that our inequality is unique and tight, with the exclusion of the regime in which $p_d \geqslant \tilde{p}$, for which the method returns the trivial inequality $p_- \leqslant \tilde{p}$, which follows immediately from Eq. (B1). As we will see, in actual experiments one has $p_d \ll \tilde{p}$.

We can now prove the theorems by obtaining specific values for $\tilde{p}$ using noncontextuality and the operational equivalence of condition 1 of each theorem.

Proof (proof of Theorem 1). By Eq. (12) and measurement noncontextuality we have

$$p_{M_W}(x|\lambda) = q(x-1)p_{M_\mathcal{E}}(y=1|\lambda) + q(x)p_{M_\mathcal{E}}(y=0|\lambda). \tag{B7}$$

Since the median of $q(x)$ is 0 we have $\int_{-\infty}^{0} q(x-1)dx \leqslant \int_{-\infty}^{0} q(x)dx = \frac{1}{2}$. In any ontological model, $\sum_y p_{M_\mathcal{E}}(y|\lambda) = 1$ for every λ. Integrating both sides of Eq. (B7) from $-\infty$ to 0 then gives Eq. (B1) with $\tilde{p} = \frac{1}{2}$. Hence, we can apply Lemma 1 to obtain the result. ■

Proof (proof of Theorem 2). By Eq. (23) and measurement noncontextuality we have

$$p_{M_W}(\gamma|\lambda) = p_{M_{\text{triv}}}(\gamma|\lambda). \tag{B8}$$

By definition $p_{M_{\text{triv}}}(\gamma|\lambda)$ is independent of λ and $\int_{-\infty}^{0} p_{M_{\text{triv}}}(\gamma|\lambda)d\gamma = \frac{1}{2}$. Integrating both sides of Eq. (B8) from $-\infty$ to 0 then gives Eq. (B1) with $\tilde{p} = \frac{1}{2}$. Hence, we can apply Lemma 1 to obtain the result. ■

Proof (proof of Theorem 3). By Eq. (29) and measurement noncontextuality we have

$$p_{M_W}(x|\lambda) = p_m p_{M_\mathcal{E}}(x|\lambda) + (1-p_m)p_{M_{\text{triv}}}(x|\lambda). \tag{B9}$$

By definition $p_{M_{\text{triv}}}(x|\lambda)$ is independent of λ and $\int_{-\infty}^{0} p_{M_{\text{triv}}}(x|\lambda)dx = \frac{1}{2}$. In any ontological model, $\int_{-\infty}^{0} p_{M_\mathcal{E}}(x|\lambda)dx \leqslant \int_{-\infty}^{\infty} p_{M_\mathcal{E}}(x|\lambda)dx = 1$. Integrating both sides of (B9) from $-\infty$ to 0 then gives Eq. (B1) with $\tilde{p} = p_m + (1-p_m)\frac{1}{2} = \frac{1+p_m}{2}$. Applying Lemma 1 gives the result. ■

Notice that the tightness of the inequality proven in Lemma 1 does not automatically imply that the inequalities in Theorems 1–3 are tight, because Eqs. (B7)–(B9) are stronger constraints than Eq. (B1) with the relevant value of $\tilde{p}$. Since Eqs. (B7) and (B8) reflect an infinite number of operational equivalences (one for each value of x), for Theorems 1 and 2 this issue cannot be straightforwardly settled using the techniques from [14] alone because those only apply to finite sets of equivalences. It may be possible to gain some confidence by using a series of increasingly fine-grained but nevertheless finite operational equivalences. Theorem 3 is a somewhat easier case: since it is intended to apply to a finite number of outcomes, for each number of outcomes there will in fact be a finite set of equivalences for which the relevant polytope could be calculated. In this work we leave the tightness of the inequalities in Theorems 1–3 as open problems, but we find the tightness of the inequality in Lemma 1 quite suggestive.

APPENDIX C: PROOF OF THEOREM 4 (PLUS EXTENSION TO IMAGINARY WEAK VALUES AND FINITE VERSION)

We will use the same structure as in Appendix B above, with the main argument in the form of a lemma.

Lemma 2 (noncontextuality inequality template 2). Suppose we have a noncontextual ontological model and that

(1) for any input λ, the probability of a negative outcome of $[x|M_W]$ is bounded by some value independent of the ontic state:

$$\int_{-\infty}^{0} p_{M_W}(x|\lambda)dx \leqslant \tilde{p}; \tag{C1}$$

(2) given the sequential measurement $[x, y|M_F \circ M_W]$, define $[y|\tilde{M}_F] := \int dx[x, y|M_F \circ M_W]$; then there exists $p_d \in [0, 1]$ such that

$$[y|\tilde{M}_F] \simeq (1 - p_d)[y|M_F] + p_d[y|M_D], \tag{C2}$$

for some 2-outcome measurement $[y|M_D]$;

(3) there exists an ensemble

$$\{\{q_*, P_*\}, \{q_\perp, P_\perp\}\},$$

such that

$$q_0[b = 0|S] + q_1[b = 1|S] \simeq q_* P_* + q_\perp P_\perp. \tag{C3}$$

Then if $p_- := \int_{-\infty}^0 p(x, y = 1|P_*, M_F \circ M_W)dx$ and $p_F := p(y = 1|P_*, M_F)$,

$$p_- \leqslant p_F \tilde{p} + (1 - p_F)p_d + \frac{1 - C_S}{q_*} \max\{\tilde{p} - p_d, 1 - \tilde{p}\}. \tag{C4}$$

Proof. Let us denote by $p_S(\lambda|b)$ the probability distribution associated with $[b|S]$.

From the definition of an ontological model, $C_S = \sum_{b,y\in\{0,1\}} \delta_{by} \int_\Lambda d\lambda p_{M_F}(y|\lambda) q_b p_S(\lambda|b)$. From the definition of conditional probability, $q_b p_S(\lambda|b) = p_S(\lambda)p_S(b|\lambda)$. Then

$$\begin{aligned} C_S &= \sum_{b,y\in\{0,1\}} \delta_{by} \int_\Lambda d\lambda p_{M_F}(y|\lambda) p_S(b|\lambda) p_S(\lambda) \\ &\leqslant \int_\Lambda d\lambda \max_{y\in\{0,1\}} p_{M_F}(y|\lambda) \sum_{b,y\in\{0,1\}} \delta_{by} p_S(b|\lambda) p_S(\lambda) \\ &:= \int_\Lambda d\lambda \zeta(\lambda) p_S(\lambda), \end{aligned} \tag{C5}$$

where $\zeta(\lambda) := \max_{y\in\{0,1\}} p_{M_F}(y|\lambda)$. We now work out some inequalities that we need in order to bound p_-. Let us now split the set of ontological variables Λ in the union of two disjoint sets: $\Lambda = \Lambda_0 \sqcup \Lambda_1$,

$$\begin{aligned} \Lambda_0 &= \{\lambda \in \Lambda | p_{M_F}(y = 0|\lambda) \geqslant p_{M_F}(y = 1|\lambda)\}, \\ \Lambda_1 &= \{\lambda \in \Lambda | p_{M_F}(y = 1|\lambda) > p_{M_F}(y = 0|\lambda)\}. \end{aligned}$$

Note that Λ_0 (Λ_1) is the set of λ's that are more likely than not to fail (pass) the postselection measurement.

Inequality 1. For every $\lambda \in \Lambda$,

$$\int_{-\infty}^0 dx p_{M_F \circ M_W}(x, y = 1|\lambda) \leqslant \int_{-\infty}^0 dx p_{M_W}(x|\lambda) \leqslant \tilde{p}, \tag{C6}$$

where we have used (C1).

Inequality 2. For every $\lambda \in \Lambda_0$,

$$\begin{aligned} \int_{-\infty}^0 dx p_{M_F \circ M_W}(x, y = 1|\lambda) &\leqslant \int_{-\infty}^{+\infty} dx p_{M_F \circ M_W}(x, y = 1|\lambda) = (1 - p_d)p_{M_F}(y = 1|\lambda) + p_d p_{M_D}(y = 1|\lambda) \\ &= (1 - p_d)[1 - \zeta(\lambda)] + p_d p_{M_D}(y = 1|\lambda) \leqslant (1 - p_d)[1 - \zeta(\lambda)] + p_d, \end{aligned} \tag{C7}$$

where we used measurement noncontextuality applied to the operational equivalence of Eq. (C2) and the definition of $\zeta(\lambda)$ in Λ_0.

We can now use these inequalities to give an upper bound to p_-. We are going to use Eq. (C6) for $\lambda \in \Lambda_1$ and Eq. (C7) for $\lambda \in \Lambda_0$:

$$\begin{aligned} p_- &= \int_{-\infty}^0 dx p(x, y = 1|P_*, M_F \circ M_W) = \sum_{i=0}^1 \int_{-\infty}^0 dx \int_{\Lambda_i} d\lambda p_{P_*}(\lambda) p_{M_F \circ M_W}(x, y = 1|\lambda) \\ &\leqslant (1 - p_d) \int_{\Lambda_0} d\lambda p_{P_*}(\lambda)[1 - \zeta(\lambda)] + p_d \int_{\Lambda_0} d\lambda p_{P_*}(\lambda) + \tilde{p} \int_{\Lambda_1} d\lambda p_{P_*}(\lambda). \end{aligned} \tag{C8}$$

Let us analyze the various terms separately:

$$\begin{aligned} \int_{\Lambda_0} d\lambda p_{P_*}(\lambda) &= \int_{\Lambda_0} d\lambda p_{P_*}(\lambda) p_{M_F}(y = 0|\lambda) + \int_{\Lambda_1} d\lambda p_{P_*}(\lambda) p_{M_F}(y = 0|\lambda) \\ &\quad + \int_{\Lambda_0} d\lambda p_{P_*}(\lambda)[1 - p_{M_F}(y = 0|\lambda)] - \int_{\Lambda_1} d\lambda p_{P_*}(\lambda) p_{M_F}(y = 0|\lambda) \end{aligned} \tag{C9}$$

$$= 1 - p_F + \int_{\Lambda_0} d\lambda p_{P_*}(\lambda)[1 - \zeta(\lambda)] - \int_{\Lambda_1} d\lambda p_{P_*}(\lambda)[1 - \zeta(\lambda)], \tag{C10}$$

where we used $\int_\Lambda d\lambda p_{P_*}(\lambda) p_{M_F}(y = 0|\lambda) = 1 - p_F$ and $p_{M_F}(y = 0|\lambda) = \zeta(\lambda)$ for $\lambda \in \Lambda_0$ and $p_{M_F}(y = 0|\lambda) = 1 - \zeta(\lambda)$ for $\lambda \in \Lambda_1$. Similarly,

$$\int_{\Lambda_1} d\lambda p_{P_*}(\lambda) = p_F - \int_{\Lambda_0} d\lambda p_{P_*}(\lambda)[1 - \zeta(\lambda)] + \int_{\Lambda_1} d\lambda p_{P_*}(\lambda)[1 - \zeta(\lambda)]. \tag{C11}$$

Substituting these in Eq. (C8) we find

$$p_- \leqslant \tilde{p}p_F + (1-p_F)p_d + (1-\tilde{p})\int_{\Lambda_0} d\lambda p_{P_*}(\lambda)[1-\zeta(\lambda)] + (\tilde{p}-p_d)\int_{\Lambda_1} d\lambda p_{P_*}(\lambda)[1-\zeta(\lambda)]$$
$$\leqslant \tilde{p}p_F + (1-p_F)p_d + \max\{\tilde{p}-p_d, 1-\tilde{p}\}\int_{\Lambda} d\lambda p_{P_*}(\lambda)[1-\zeta(\lambda)]. \quad \text{(C12)}$$

By preparation noncontextuality, Eq. (C3) implies $p_S(\lambda) = q_* p_{P_*}(\lambda) + q_\perp p_{P_\perp}(\lambda) \geqslant q_* p_{P_*}(\lambda)$. Combining this with Eq. (C5), we have

$$1 - C_S = \int_\Lambda d\lambda p_S(\lambda)[1-\zeta(\lambda)] \geqslant q_* \int_\Lambda d\lambda p_{P_*}(\lambda)[1-\zeta(\lambda)]. \quad \text{(C13)}$$

Substituting in the previous equation, we obtain the claimed bound. ■

Concerning tightness, we used the same approach as for Lemma 1, fixing the numerical values for $\tilde{p}$, p_d, q_*, q_0. For relevant choices of parameters we observe that the inequality defines a facet in the "noncontextuality polytope." Furthermore, we provide numerical tools to derive all noncontextual inequalities for all choices of parameters; see Appendix H.

Proof (proof of Theorem 4). By Eq. (34) and measurement noncontextuality we have

$$p_{M_W}(x|\lambda) = q(x-1)p_{M_\varepsilon}(y=1|\lambda) + q(x)p_{M_\varepsilon}(y=0|\lambda). \quad \text{(C14)}$$

Since the median of $q(x)$ is 0 we have $\int_{-\infty}^{0} q(x-1)dx \leqslant \int_{-\infty}^{0} q(x)dx = \frac{1}{2}$. In any ontological model, $\sum_y p_{M_\varepsilon}(y|\lambda) = 1$ for every λ. Integrating both sides of Eq. (C14) from $-\infty$ to 0 then gives Eq. (C1) with $\tilde{p} = \frac{1}{2}$. Noting that $\tilde{p} = \frac{1}{2}$ gives

$$\max\{\tilde{p}-p_d, 1-\tilde{p}\} = \max\left\{\frac{1}{2}-p_d, \frac{1}{2}\right\} = \frac{1}{2}, \quad \text{(C15)}$$

we can apply Lemma 2 to obtain the result. ■

The extensions to imaginary weak values and to finite versions can be easily derived from Lemma 2 following the same procedure as at the end of Appendix B. The situation regarding tightness of the inequalities also mirrors the discussion there.

APPENDIX D: NOISY IMPLEMENTATION OF THE WEAK VALUE

Lemma 3. In quantum theory, a weak measurement of the projector $\mathcal{E}$ with initial spread of the pointer s and imperfect postselection of Π_ϕ with ϵ-unbiased noise as in Eq. (17) achieves

$$p_-^{\text{noisy}} = \frac{p_F}{2} - \frac{1}{\sqrt{\pi}s}\text{Re}(\langle[y=1|M_F]\mathcal{E}\rangle_{\rho_*}) + o\left(\frac{1}{s}\right), \quad \text{(D1)}$$

where $C_S = 1-\epsilon$. The operational equivalences required by Theorem 1 are satisfied, and those of Theorem 4 can be satisfied by introducing the preparations ($d \equiv \text{Tr}[\mathbb{1}]$)

$$\sigma_0 = \frac{\mathbb{1}-\Pi_\phi}{d - \text{Tr}\,\Pi_\phi}, \quad \sigma_1 = \frac{\Pi_\phi}{\text{Tr}\,\Pi_\phi}. \quad \text{(D2)}$$

Remarks. Note that the preparations $[b|S]$ were taken to have singular density operators, but this assumption does not imply an extra idealization. In fact, if we add unbiased noise to S, $\sigma_1 = (1-\delta)\frac{\Pi_\phi}{\text{Tr}\,\Pi_\phi} + \delta\frac{\mathbb{1}}{d}$ and similarly for σ_0, we could absorb δ by a redefinition of ϵ. Also note that exactly the same proof shows that the operational equivalences required by Theorems 2 and 3, as well as for the imaginary weak values and finite versions of Theorem 4, do hold. Finally, for the imaginary weak value version, p_- has an expression similar to Eq. (D1), but involving the imaginary part of the weak value.

Proof. The weak measurement scheme with ϵ-unbiased noise in the postselection coincides with the standard scheme described in Sec. II B with the only difference that the postselection is taken to be

$$\{[y=1|M_F], [y=0|M_F]\} = (1-2\epsilon)\{\Pi_\phi, \mathbb{1}-\Pi_\phi\} + 2\epsilon\{\mathbb{1}/2, \mathbb{1}/2\}.$$

Concerning the relation between C_S and ϵ:

$$\begin{aligned} C_S &= q_0 p(y=0|b=0,S,M_F) + q_1 p(y=1|b=1,S,M_F) \\ &= q_0[(1-2\epsilon)p(y=0|b=0,S,\{\Pi_\phi,\mathbb{1}-\Pi_\phi\}) + \epsilon] \\ &\quad + q_1[(1-2\epsilon)p(y=1|b=1,S,\{\Pi_\phi,\mathbb{1}-\Pi_\phi\}) + \epsilon] \\ &= q_0(1-\epsilon) + q_1(1-\epsilon) = 1-\epsilon. \end{aligned} \quad \text{(D3)}$$

Operational equivalences. The operational equivalences of Theorem 1 are satisfied by following the same argument as described in the main text for the ideal case, since none of them involve the postselection.

Concerning the equivalences required for Theorem 4 and related imaginary or finite versions, the ones that do not follow immediately from previous arguments are Eq. (35) and Eq. (36).

To prove Eq. (35) we can start with the definition

$$\begin{aligned} [y|\tilde{M}_F] &:= \int_{-\infty}^{+\infty} [x, y|M_F \circ M_W] \\ &= \int_{-\infty}^{+\infty} dx N_x^\dagger [y|M_F] N_x = \mathcal{M}^\dagger([y|M_F]), \end{aligned} \quad \text{(D4)}$$

and, using Eq. (7), obtain

$$\begin{aligned} [y=1|\tilde{M}_F] &= \mathcal{M}^\dagger([y=1|M_F]) \\ &= (1-2\epsilon)\mathcal{M}^\dagger(\Pi_\phi) + \epsilon\mathbb{1} = p_d[y=1|M_F] \\ &\quad + (1-p_d)[(1-2\epsilon)\mathcal{M}_D^\dagger(\Pi_\phi) + \epsilon\mathbb{1}]. \end{aligned} \quad \text{(D5)}$$

By defining a POVM $\{M_D, \mathbb{1}-M_D\}$ with $M_D = (1-2\epsilon)\mathcal{M}_D^\dagger(\Pi_\phi) + \epsilon\mathbb{1}$, we can see that Eq. (35) is satisfied with the same p_d as in the ideal case, $p_d = (1-e^{-1/4s^2})/2$.

Moving on to Eq. (36), to satisfy it we need a careful choice of $P_\perp$ with the aim of maximizing q_* and hence the violation. We will leave q_* as a free parameter, but note that a choice satisfying Eq. (36) always exists for any choice of P_* given by ρ_*:

$$q_* = 1/d, \quad \rho_\perp = \frac{\mathbb{1} - \rho_*}{d-1}, \quad q_1 = \frac{\mathrm{Tr}[\Pi_\phi]}{d}.$$

In fact, with these choices,

$$q_*\rho_* + q_\perp\rho_\perp = q_0\sigma_0 + q_1\sigma_1 = \mathbb{1}/d. \tag{D6}$$

Expression for p_-^{noisy}. For both the definition of p_- of Theorem 1 and that of Theorem 4, using Eq. (4),

$$p_-^{\text{noisy}} = \epsilon \int_{-\infty}^{0} dx\, \mathrm{Tr}(N_x^\dagger N_x \rho_*) + (1-2\epsilon)\int_{-\infty}^{0} \mathrm{Tr}(N_x^\dagger \Pi_\phi N_x \rho_*) dx. \tag{D7}$$

For the first term, since $N_x^\dagger N_x = G_s^2(x-1)\mathcal{E} + G_s^2(x)\mathcal{E}^\perp$, using the integral $\int_{-\infty}^0 dx G_s^2(x-1) = \frac{1}{2}\,\mathrm{erfc}\,(\frac{1}{s})$ expressed using the complementary error function $\mathrm{erfc}(x) \equiv 1 - \mathrm{erf}(x) \equiv 1 - \frac{1}{\sqrt{\pi}}\int_{-x}^{x} e^{-t^2}dt$ and the expansion $\mathrm{erfc}(1/s) = 1 - 2/(\sqrt{\pi}s) + o(1/s)$,

$$\int_{-\infty}^{0} dx\, \mathrm{Tr}(N_x^\dagger N_x \rho_*) = \frac{1}{2}\,\mathrm{erfc}\left(\frac{1}{s}\right)p_\mathcal{E} + \frac{1}{2}(1-p_\mathcal{E}) = \frac{1}{2} - \frac{p_\mathcal{E}}{\sqrt{\pi}s} + o\left(\frac{1}{s}\right), \tag{D8}$$

where $p_\mathcal{E} = \mathrm{Tr}(\mathcal{E}\rho_*)$.

For the second term, from Eq. (4) and the integral $\int_{-\infty}^0 G_s(x-1)G_s(x)dx = \frac{e^{-1/4s^2}}{2}\,\mathrm{erfc}\,(\frac{1}{2s})$ we get

$$\begin{aligned}
&\int_{-\infty}^{0} \mathrm{Tr}(N_x^{s\dagger}\Pi_\phi N_x \rho_*)dx \\
&= \frac{1}{2}\,\mathrm{erfc}\left(\frac{1}{s}\right)\mathrm{Tr}(\mathcal{E}\Pi_\phi\mathcal{E}\rho_*) + \frac{e^{-1/(4s^2)}}{2}\,\mathrm{erfc}\left(\frac{1}{2s}\right) \\
&\quad\times \mathrm{Tr}[(\mathcal{E}^\perp\Pi_\phi\mathcal{E} + \mathcal{E}\Pi_\phi\mathcal{E}^\perp)\rho_*] + \frac{1}{2}\mathrm{Tr}(\mathcal{E}^\perp\Pi_\phi\mathcal{E}^\perp\rho_*) \\
&= \frac{1}{2}\mathrm{Tr}(\Pi_\phi\rho_*) - \frac{1}{2\sqrt{\pi}s}\mathrm{Tr}[(\Pi_\phi\mathcal{E} + \mathcal{E}\Pi_\phi)\rho_*] + o\left(\frac{1}{s}\right),
\end{aligned} \tag{D9}$$

and we note that $\mathrm{Tr}\,[(\Pi_\phi\mathcal{E} + \mathcal{E}\Pi_\phi)\rho_*] = 2\,\mathrm{Re}(\langle\Pi_\phi\mathcal{E}\rangle_{\rho_*})$. Substituting everything into the expression for p_-^{noisy},

$$p_-^{\text{noisy}} = \frac{p_F}{2} - \frac{1}{\sqrt{\pi}s}\mathrm{Re}(\langle[y=1|M_F]\mathcal{E}\rangle_{\rho_*}) + o\left(\frac{1}{s}\right). \tag{D10}$$

■

APPENDIX E: MEASUREMENTS OF POINTER MOMENTUM

Now we calculate p_-, the probability of a negative value of p under the postselection. For simplicity we will only consider the ideal case, where $[y=1|M_F]$ is a projection Π_ϕ. However, the noisy case can be derived extending the treatment below in the same way as we did with the position measurement of the pointer in Appendix D. Thus,

$$\begin{aligned}
p_- &= \int_{-\infty}^{0} \mathrm{Tr}(N_\gamma^\dagger \Pi_\phi N_\gamma \rho_*)d\gamma \\
&= \frac{1}{2}[\mathrm{Tr}(\mathcal{E}\Pi_\phi\mathcal{E}\rho_*) + \mathrm{Tr}(\mathcal{E}^\perp\Pi_\phi\mathcal{E}^\perp\rho_*) \\
&\quad + \alpha\,\mathrm{Tr}(\mathcal{E}^\perp\Pi_\phi\mathcal{E}\rho_*) + \alpha^*\,\mathrm{Tr}(\mathcal{E}\Pi_\phi\mathcal{E}^\perp\rho_*)]
\end{aligned} \tag{E1}$$

with integral [recalling Eq. (2)]

$$\begin{aligned}
\alpha &= 2\int_{-\infty}^{0} |\langle\gamma|\Psi\rangle_P|^2 \exp(-i\gamma)d\gamma \\
&= \exp\left(-\frac{1}{4s^2}\right)\left[1 + \mathrm{erf}\left(\frac{i}{2s}\right)\right].
\end{aligned} \tag{E2}$$

To calculate α^* recall that the erf of a purely imaginary number is purely imaginary. Using $\alpha \approx 1 + \frac{i}{\sqrt{\pi}s}$ and $\mathrm{Tr}[(\mathcal{E}+\mathcal{E}^\perp)\Pi_\phi(\mathcal{E}+\mathcal{E}^\perp)\rho_*] = \mathrm{Tr}(\Pi_\phi\rho_*) = p_F$ we find, at leading order in $1/s$,

$$p_- \approx \frac{p_F}{2} + \frac{1}{\sqrt{\pi}s}\mathrm{Re}[i\,\mathrm{Tr}(\mathcal{E}^\perp\Pi_\phi\mathcal{E}\rho_*)]. \tag{E3}$$

Since $\mathcal{E}^\perp = \mathbb{1} - \mathcal{E}$ and $\mathrm{Im}[\mathrm{Tr}(\mathcal{E}\Pi_\phi\mathcal{E}\rho_*)] = 0$ this gives, at leading order in $1/s$,

$$p_- \approx \frac{1}{2} - \frac{\mathrm{Im}(\langle\Pi_\phi\mathcal{E}\rangle_{\rho_*})}{\sqrt{\pi}s}. \tag{E4}$$

APPENDIX F: QUBIT POINTERS

In Ref. [23] weak measurements using qubit pointers are constructed, with the weakness controlled by a parameter in the interaction between the system and pointer. It turns out that, as in the continuous-pointer case, one can also use a fixed interaction and control the weakness using a parameter in the pointer state. For consistency we take that approach here.

The interaction we consider is $U = \mathcal{E}\otimes Z + \mathcal{E}^\perp\otimes\mathbb{1}$ where Z denotes the Pauli-Z operator on the qubit pointer. This interaction is basically a controlled-phase gate where the control is $\mathcal{E}$ versus $\mathcal{E}^\perp$. Indeed, by preparing the pointer in $|X=-1\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$ and measuring Pauli X on the pointer one can carry out a strong measurement of $\mathcal{E}$ with the usual disturbance. On the other hand, since $Z|0\rangle = |0\rangle$, preparing the pointer in $|0\rangle$ would mean U acts as identity on the system and hence causes no disturbance. This suggests we can achieve a weak measurement by taking an initial pointer state of $|\Psi_\epsilon\rangle = \cos\epsilon\,|0\rangle - \sin\epsilon\,|1\rangle$, where ϵ is small. Measuring X on the pointer gives Kraus operators [here $|X=1\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$]

$$\begin{aligned}
N_{\pm1} &= \langle X = \pm1|U|\Psi_\epsilon\rangle \\
&= \frac{1}{\sqrt{2}}[\mathcal{E}(\cos\epsilon \pm \sin\epsilon) + \mathcal{E}^\perp(\cos\epsilon \mp \sin\epsilon)] \\
&= \frac{1}{\sqrt{2}}[\cos\epsilon\,\mathbb{1} \pm \sin\epsilon(\mathcal{E} - \mathcal{E}^\perp)],
\end{aligned} \tag{F1}$$

and hence POVM elements

$$N_{\pm1}^\dagger N_{\pm1} = \tfrac{\mathbb{1}}{2} \pm \cos\epsilon\sin\epsilon(\mathcal{E} - \mathcal{E}^\perp), \tag{F2}$$

so that

$$N_{+1}^\dagger N_{+1} = (1-p_m)\tfrac{\mathbb{1}}{2} + p_m \mathcal{E},$$
$$N_{-1}^\dagger N_{-1} = (1-p_m)\tfrac{\mathbb{1}}{2} + p_m \mathcal{E}^\perp, \tag{F3}$$

where $p_m = 2\cos\epsilon\sin\epsilon = \sin(2\epsilon)$. Hence, Eq. (29) is satisfied.

If we ignore the outcome of the measurement on the pointer and then apply a channel

$$\begin{aligned}\mathcal{M}(\rho) &= N_{+1}\rho N_{+1}^\dagger + N_{-1}\rho N_{-1}^\dagger \\ &= \cos^2\epsilon\,\rho + \sin^2\epsilon(\mathcal{E}-\mathcal{E}^\perp)\rho(\mathcal{E}-\mathcal{E}^\perp) \\ &= (1-p_d)\rho + p_d\mathcal{M}^D(\rho),\end{aligned} \tag{F4}$$

where $p_d = \sin^2\epsilon$ and $\mathcal{M}^D(\rho) = (\mathcal{E}-\mathcal{E}^\perp)\rho(\mathcal{E}-\mathcal{E}^\perp)$, Eq. (30) is satisfied.

Finally, considering a perfect postselection onto a projector Π_ϕ, we can calculate

$$\begin{aligned}p_- = \mathrm{Tr}(N_{-1}^\dagger \Pi_\phi N_{-1}\rho_*) &= \tfrac{1}{2}\{\cos^2\epsilon\,\mathrm{Tr}(\Pi_\phi\rho_*) \\ &+ \sin^2\epsilon\,\mathrm{Tr}[(\mathcal{E}-\mathcal{E}^\perp)\Pi_\phi(\mathcal{E}-\mathcal{E}^\perp)\rho_*] \\ &- \sin\epsilon\cos\epsilon\,\mathrm{Tr}[(\mathcal{E}-\mathcal{E}^\perp)\Pi_\phi\rho_* + \Pi_\phi(\mathcal{E}-\mathcal{E}^\perp)\rho_*]\}.\end{aligned} \tag{F5}$$

Expanding to first order in ϵ gives

$$\begin{aligned}p_- &\approx \frac{p_F}{2} - \frac{\epsilon}{2}\mathrm{Tr}[(\mathcal{E}-\mathcal{E}^\perp)\Pi_\phi\rho_* + \Pi_\phi(\mathcal{E}-\mathcal{E}^\perp)\rho_*] \\ &= \frac{p_F}{2} - \epsilon\,\mathrm{Re}\{\mathrm{Tr}[\Pi_\phi(\mathcal{E}-\mathcal{E}^\perp)\rho_*]\} \\ &= \frac{p_F}{2} - \epsilon\{2\,\mathrm{Re}[\mathrm{Tr}(\Pi_\phi\mathcal{E}\rho_*)] - p_F\},\end{aligned} \tag{F6}$$

and since $p_m \approx 2\epsilon$ we obtain, at leading order in ϵ,

$$p_- \approx p_F\frac{1+p_m}{2} - 2\epsilon\,\mathrm{Re}(\langle\Pi_\phi\mathcal{E}\rangle_{\rho_*}). \tag{F7}$$

APPENDIX G: DETAILS OF MINIMAL-DISTURBANCE ONTOLOGICAL MODEL

The weak measurement $\mathcal{M}^W$ disturbs the system so that the operational probabilities for the postselection following it, $p(y|P_*,\mathcal{M},M_F)$, differ from those that would be obtained without the weak measurement, $p(y|P_*,M_F)$. Normally the postselection becomes slightly more likely, i.e., $\epsilon := p(1|P_*,\mathcal{M},M_F) - p(1|P_*,M_F) > 0$, because the postselection is chosen almost orthogonal to the preparation and the weak measurement makes the state of the system slightly mixed. We will construct a model under this assumption, but if the opposite is true then we simply need to exchange the roles of $y=0$ and $y=1$ in the rest of the discussion. By normalization $\epsilon = p(0|P_*,M_F) - p(0|P_*,\mathcal{M},M_F)$, and clearly $\epsilon \leqslant 1$ [indeed ϵ is just the total variation distance between $p(y|P_*,M_F)$ and $p(y|P_*,\mathcal{M},M_F)$]. Hence we can define

$$D(y'|y) = \delta_{y'y} + \frac{\epsilon}{p(0|P_*,M_F)}S(y'|y). \tag{G1}$$

$$S(y'|y) = \begin{cases} -1, & y=0, y'=0, \\ 1, & y=0, y'=1, \\ 0, & y=1. \end{cases} \tag{G2}$$

This is a "minimally disturbing" [29] conditional distribution such that

$$p(y'|P_*,\mathcal{M},M_F) = \sum_y D(y'|y)p(y|P_*,M_F). \tag{G3}$$

We use this disturbance in the representation of $\mathcal{M}$ in the ontological model:

$$p_{\mathcal{M}^W}(x,\lambda'|\lambda) = p(x|P_*,b=1,\mathcal{M},M_F,y=\lambda')D(y'=\lambda'|y=\lambda). \tag{G4}$$

By construction

$$p_{\mathcal{M}}(\lambda'|\lambda) = \int_{-\infty}^{\infty} p_{\mathcal{M}^W}(x,\lambda'|\lambda)dx = D(y'=\lambda'|y=\lambda), \tag{G5}$$

and we have that

$$D(y'|y) = (1-p_d)\delta_{y'y} + p_d\left[\delta_{y'y} + \frac{\epsilon}{p(0|P_*,M_F)p_d}S(y'|y)\right], \tag{G6}$$

which suggests that in order to satisfy condition 2 of Theorems 1–3 we should set

$$p_{\mathcal{M}^D}(\lambda'|\lambda) = \delta_{\lambda'\lambda} + \frac{\epsilon}{p(0|P_*,M_F)p_d}S(y'=\lambda'|y=\lambda). \tag{G7}$$

It is easy to see that this is normalized and is clearly positive except perhaps for

$$p_{\mathcal{M}^D}(\lambda'=0|\lambda=0) = 1 - \frac{\epsilon}{p(0|P_*,M_F)p_d}, \tag{G8}$$

which is positive provided $p_d \geqslant \frac{\epsilon}{p(0|P_*,M_F)}$. To check this we note that the operational equivalence of condition 2 on $\tilde{M}_F$ tells us that

$$\begin{aligned}p(1|P_*,\mathcal{M},M_F) &= (1-p_d)p(1|P_*,M_F) \\ &+ p_d\,p(y=1|P_*,\mathcal{M}^D,M_F)\end{aligned} \tag{G9}$$

so that, since $p(y=1|P_*,\mathcal{M}^D,M_F) \leqslant 1$,

$$\begin{aligned}\frac{\epsilon}{p(0|P_*,M_F)} &= \frac{p(1|P_*,\mathcal{M},M_F) - p(1|P_*,M_F)}{1-p(1|P_*,M_F)} \\ &= p_d\frac{p(y=1|P_*,\mathcal{M}^D,M_F) - p(1|P_*,M_F)}{1-p(1|P_*,M_F)} \leqslant p_d,\end{aligned} \tag{G10}$$

as required.

APPENDIX H: ALGORITHMIC APPROACH TO TIGHTNESS

We discretize the problem and use the algorithmic approach of Ref. [14], to which we refer for extra details, in order to verify that the noncontextuality inequalities of Lemmas 1 and 2 are indeed facet inequalities of the noncontextuality polytope describing the relevant statistics. We first set up the general algorithmic problem and then see how to apply it to each theorem.

1. Setting up the problem

Since we will be dealing with arrays of procedures it will be useful to number them as follows:

$$P_\perp \leftrightarrow P_1, \quad P_* \leftrightarrow P_2, \quad [b=0|S] \leftrightarrow P_3, \quad [b=1|S] \leftrightarrow P_4. \tag{H1}$$

The operational equivalence of Eq. (C3) can thus be written as

$$q_\perp P_1 + (1-q_\perp)P_2 \simeq q_0 P_3 + (1-q_0)P_4. \tag{H2}$$

Since the definition of p_- and the relevant constraints only involve a coarse graining of the measurement outcome of M_W (the weak measurement), we denote a binary-outcome coarse graining of M_W as

$$\begin{aligned}\left[X=-1\middle|M_W^{\text{bin}}\right] &= \int_{-\infty}^{0} dx[x|M_W],\\ \left[X=+1\middle|M_W^{\text{bin}}\right] &= \int_{0}^{\infty} dx[x|M_W].\end{aligned} \tag{H3}$$

Henceforth, we will consider the sequential measurement $M_F \circ M_W^{\text{bin}}$ rather than $M_F \circ M_W$. The operational equivalence of Eq. (C2) used in Lemma 2 is

$$\begin{aligned}[y|\tilde{M}_F] &= \sum_{X=\pm1}\left[X, y\middle|M_F \circ M_W^{\text{bin}}\right]\\ &\simeq (1-p_d)[y|M_F] + p_d[y|M_D].\end{aligned} \tag{H4}$$

Finally Eq. (C1) (which appears in both lemmas) becomes the condition

$$p_{M_W^{\text{bin}}}(X=-1|\lambda) \leqslant \tilde{p} \in [0,1] \quad \forall \lambda \in \Lambda. \tag{H5}$$

Similarly we number the relevant measurements as $\{M_1, M_2, M_3\}$ and their outcomes by $m \in \{1,2,3,4\}$, defining events $[m|M_i]$ as

$$\begin{aligned}M_1 &: [1|M_1] = [1|M_F], \quad [2|M_1] = [0|M_F],\\ M_2 &: [1|M_2] = [1|M_D], \quad [2|M_2] = [0|M_D],\\ M_3 &: [1|M_3] = [X=-1, y=1|M],\\ &\quad [2|M_3] = [X=-1, y=0|M],\\ &\quad [3|M_3] = [X=+1, y=1|M],\\ &\quad [4|M_3] = [X=+1, y=0|M].\end{aligned} \tag{H6}$$

The operational equivalence of Eq. (H4) can then be restated as

$$\begin{aligned}(1-p_d)[1|M_1] + p_d[1|M_2] &\simeq [1|M_3] + [3|M_3],\\ (1-p_d)[2|M_1] + p_d[2|M_2] &\simeq [2|M_3] + [4|M_3],\end{aligned} \tag{H7}$$

while Eq. (H5) becomes

$$p_{M_3}(1|\lambda) + p_{M_3}(2|\lambda) \leqslant \tilde{p} \in [0,1]. \tag{H8}$$

Applying measurement noncontextuality to Eq. (H7) gives two linear constraints on the P_{M_i}, on top of which we have (H8), normalization, and positivity.

For any fixed λ, we can see an assignment of the $p_{M_i}(m|\lambda)$ as an 8-component vector. The set of all assignments compatible with the above constraints defines a polytope in this space, which we denote as *weakvaluespolysymbN* in the accompanying code [30]. Its vertices will be denoted by κ. The vertex assignments in the polytope are denoted by $p_{M_i}(m|\kappa)$. For every λ, we can decompose $p_{M_i}(m|\lambda)$ as

$$p_{M_i}(m|\lambda) = \sum_\kappa w(\kappa|\lambda) p_{M_i}(m|\kappa), \tag{H9}$$

where $w(\kappa|\lambda) \geqslant 0$, $\sum_\kappa w(\kappa|\lambda) = 1$. Hence, we can characterize all possible assignments by computing the vertex assignments. Doing the vertex enumeration with SageMath we find there are 16 such vertices, $\kappa_1, \ldots, \kappa_{16}$.

2. Tightness of the inequality in Lemma 1

Let us consider inequality in Lemma 1. Reference [14] does not consider transformation noncontextuality, and it is not obvious how to extend the approach there to transformation noncontextuality in general. But for checking tightness in our scenario it happens that we do not require such an extension. We will prove the following result, showing that a transformation and measurement noncontextual model for the weak-value experiment exists if there exists a model satisfying the original assumptions of Ref. [9]—i.e., measurement noncontextuality and outcome determinism.

Lemma 4. Suppose there exists a given model which satisfies Eq. (34), is measurement noncontextual for the equivalence Eq. (35), and represents M_F for all λ with $p_{M_F}(y|\lambda) \in \{0,1\}$. Then there exists a derived model which satisfies Eq. (B1), is transformation noncontextual for the equivalence Eq. (B2), and makes the same operational predictions as the given model.

Proof. The derived model, probabilities of which we denote using $\mathfrak{p}$, will take the ontic state λ to be a determination of y (as in Sec. V A). In fact, it is constructed from the given model, which we denote by the usual p, by coarse-graining together all ontic states that assign the same outcome y to M_F. In particular we set

$$\mathfrak{p}_{P_*}(\lambda = y) := \int_\Lambda p_{M_F}(y|\lambda) p_{P_*}(\lambda) d\lambda, \tag{H10}$$

so that we have the same predictions for an immediate measurement of M_F. We also set

$$\begin{aligned}&\mathfrak{p}_{\mathcal{M}^W}(x, \lambda' = y'|\lambda = y)\\ &:= \frac{1}{\mathfrak{p}_{P_*}(\lambda = y)} \int_\Lambda p_{M_F \circ M_W}(x, y'|\lambda) p_{M_F}(y|\lambda) p_{P_*}(\lambda) d\lambda.\end{aligned} \tag{H11}$$

This gives

$$\begin{aligned}&\sum_y \mathfrak{p}_{\mathcal{M}^W}(x, \lambda' = y'|\lambda = y)\mathfrak{p}_{P_*}(\lambda = y)\\ &\qquad = \int_\Lambda p_{M_F \circ M_W}(x, y'|\lambda) p_{P_*}(\lambda) d\lambda,\end{aligned} \tag{H12}$$

so that we also have the same predictions for $\mathcal{M}^W$ followed by M_F. From Eq. (H11), we can calculate

$$\begin{aligned}&\mathfrak{p}_{M_W}(x|\lambda = y)\\ &= \sum_{y'} \mathfrak{p}_{\mathcal{M}^W}(x, \lambda' = y'|\lambda = y)\\ &= \frac{1}{\mathfrak{p}_{P_*}(\lambda = y)} \int_\Lambda p_{M_W}(x|\lambda) p_{M_F}(y|\lambda) p_{P_*}(\lambda) d\lambda,\end{aligned} \tag{H13}$$

and hence, since the given model satisfies Eq. (34),

$$\int_{-\infty}^{0} \mathfrak{p}_{M_W}(x|\lambda = y)dx$$
$$= \frac{1}{\mathfrak{p}_{P_*}(\lambda = y)} \int_\Lambda \left[\int_{-\infty}^{0} p_{M_W}(x|\lambda)dx\right] p_{M_F}(y|\lambda)p_{P_*}(\lambda)d\lambda$$
$$\leqslant \tilde{p}\frac{1}{\mathfrak{p}_{P_*}(\lambda = y)} \int_\Lambda p_{M_F}(y|\lambda)p_{P_*}(\lambda)d\lambda = \tilde{p}, \tag{H14}$$

giving Eq. (B1) as claimed. Finally, we can calculate

$$\mathfrak{p}_{\mathcal{M}}(\lambda' = y'|\lambda = y)$$
$$= \int_{-\infty}^{\infty} \mathfrak{p}_{\mathcal{M}^W}(x, \lambda' = y'|\lambda = y)dx$$
$$= \frac{1}{\mathfrak{p}_{P_*}(\lambda = y)} \int_\Lambda p_{\tilde{M}_F}(y'|\lambda)p_{M_F}(y|\lambda)p_{P_*}(\lambda)d\lambda. \tag{H15}$$

Then since the given model is measurement noncontextual for Eq. (35) we find

$$\mathfrak{p}_{\mathcal{M}}(\lambda' = y'|\lambda = y)$$
$$= \frac{1}{\mathfrak{p}_{P_*}(\lambda = y)}\Big[(1 - p_d) \int_\Lambda p_{M_F}(y'|\lambda)p_{M_F}(y|\lambda)p_{P_*}(\lambda)d\lambda$$
$$+ p_d \int_\Lambda p_{M_D}(y'|\lambda)p_{M_F}(y|\lambda)p_{P_*}(\lambda)d\lambda\Big]$$
$$= (1 - p_d)\delta_{y'y} + p_d\mathfrak{p}_{\mathcal{M}^D}(\lambda' = y'|\lambda = y), \tag{H16}$$

where for the first term we have used outcome determinism to find $p_{M_F}(y'|\lambda)p_{M_F}(y|\lambda) = \delta_{y'y'}p_{M_F}(y|\lambda)$ and in the second we have defined

$$\mathfrak{p}_{\mathcal{M}^D}(\lambda' = y'|\lambda = y)$$
$$:= \frac{1}{\mathfrak{p}_{P_*}(\lambda = y)} \int_\Lambda p_{M_D}(y'|\lambda)p_{M_F}(y|\lambda)p_{P_*}(\lambda)d\lambda. \tag{H17}$$

Hence we satisfy transformation noncontextuality for Eq. (2). ■

We believe the converse also holds but we do not strictly require that here, since we have already proven that our inequality follows from transformation noncontextuality.

Thanks to this result we get the following algorithmic formulation for Lemma 1: consider the vertices $p_{M_i}(m|\kappa)$ from Appendix H 1 that satisfy the additional constraint $p_{M_1}(m|\kappa) \in \{0, 1\}$. To determine a set of achievable $p(m|M_i, P_k)$ we consider outcome-deterministic measurement noncontextual models given as

$$p(m|M_i, P_*) = \sum_\kappa p_*(\kappa)p_{M_i}(m|\kappa), \tag{H18}$$

where the sum is over the vertices κ satisfying the determinism constraint. Of the 16 vertices determined before, we find that 12 satisfy it and store them in 12×8 matrix *mncdetverticeswvN*.

We now project the 12 vertices down to the subspace that corresponds to the operational quantities we want to relate via noncontextuality: p_-, p_F. This subspace corresponds to the coordinates x_1 and x_5: x_1 is for the effect $[1|M_F]$ (hence related to p_F), and x_5 for $[X = -1, y = 1|M]$ (hence related to p_-). This is done by restricting the 12 vertices to the coordinates (x_1, x_5) and constructing their convex hull to yield the reduced polytope. This reduced polytope, named *mncreduceddetpolyN*, constructed in this subspace $\mathbb{R}^2$ has 4 vertices. By trying several values of p_d and $\tilde{p}$ we find they are of the form $(0, 0)$, $(0, p'_d)$, $(1, 0)$, $(1, \tilde{p})$ where $p'_d = \min\{p_d, \tilde{p}\}$, which will equal p_d for typical parameters. The H-representation of the polytope is given by $x_1, x_5 \geqslant 0$, $x_1 \leqslant 1$, and $x_1\frac{\tilde{p}-p'_d}{p'_d\tilde{p}} - \frac{x_5}{p'_d\tilde{p}} + \frac{1}{\tilde{p}} \geqslant 0$. The last inequality gives an operational constraint of $p_- \leqslant p_F\tilde{p} + (1 - p_F)p'_d = \min\{p_F\tilde{p} + (1 - p_F)p_d, \tilde{p}\}$, as expected from Lemma 1.

3. Analysis of the inequality in Lemma 2

For this case no new tricks are required and so we very closely follow [14]. If there is a measurement noncontextual model then the observed statistics $p(m|M_i, P_k)$ can be written as

$$p(m|M_i, P_k) = \sum_\kappa p_k(\kappa)p_{M_i}(m|\kappa), \tag{H19}$$

where we now sum over all 16 vertices $\{\kappa\}$. Preparation noncontextuality applied to Eq. (H2) gives

$$q_\perp p_1(\kappa) + (1 - q_\perp)p_2(\kappa) = q_0 p_3(\kappa) + (1 - q_0)p_4(\kappa), \forall\kappa. \tag{H20}$$

We thus we arrive at the following formulation for Theorem 4. In order for a noncontextual model to satisfy the assumptions of Theorem 4 and reproduce the statistics $p(m|M_i, P_j)$, the following constraints must be satisfied:

$$\forall\kappa, j : p_j(\kappa) \geqslant 0, \tag{H21}$$

$$\forall j : \sum_\kappa p_j(\kappa) = 1, \tag{H22}$$

$$\forall\kappa : q_\perp p_1(\kappa) + (1 - q_\perp)p_2(\kappa)$$
$$-[q_0 p_3(\kappa) + (1 - q_0)p_4(\kappa)] = 0, \tag{H23}$$

$$\forall i, j, m : \sum_\kappa p_{M_i}(m|\kappa)p_j(\kappa) = p(m|M_i, P_j). \tag{H24}$$

The problem can be solved by eliminating the variables $p_i(\kappa), i \in \{1, 2, 3, 4\}, \kappa \in \{\kappa_1, \kappa_2, \ldots, \kappa_{16}\}$. Since we do not care about all of the $p(m|M_i, P_j)$, for computational efficiency we first take the 16 vertices of the polytope *weakvaluespolysymbN* and cull all the coordinates from them except x_1, x_5. This is because we want to look at constraints from noncontextuality on the quantities (p_F, C_S, p_-) which are a function of these two coordinates alone.

Using the resulting set of 16 vertices projected in the (x_1, x_5) subspace, denoted *mncx1x5verticesN*, as an input to SageMath's *Polyhedron()*, we obtain a 2-dimensional 5-vertex polytope in $\mathbb{R}^2$, denoted *mncreducedpolyN*. We keep the vertices κ' of this polytope in a 5×2 matrix *redvtxN*.

The problem is now to eliminate $p_j(\kappa'), j \in \{1, 2, 3, 4\}, \kappa' \in \{\kappa'_1, \kappa'_2, \ldots, \kappa'_5\}$. Using Eq. (H23) we can manually eliminate $p_1(\kappa') = \frac{1}{q_\perp}[q_0 p_3(\kappa') + (1 -$

$q_0)p_4(\kappa') - (1-q_\perp)p_2(\kappa')]$ and arrive at the following constraints:

$$\forall \kappa', \forall j \in \{2,3,4\} : p_j(\kappa') \geqslant 0, \tag{H25}$$

$$\forall \kappa' : q_0 p_3(\kappa') + (1-q_0)p_4(\kappa') - (1-q_\perp)p_2(\kappa') \geqslant 0, \tag{H26}$$

$$\forall j \in \{2,3,4\} : \sum_{\kappa'} p_j(\kappa') = 1, \tag{H27}$$

$$\sum_{i=1}^{5} p_2(\kappa_i')\kappa_i'(0) = p(1|M_2,P_2) \equiv p_F, \tag{H28}$$

$$\sum_{i=1}^{5} p_2(\kappa_i')\kappa_i'(1) = p(1|M_4,P_2) \equiv p_-, \tag{H29}$$

$$q_0\left(\sum_{i=1}^{5} p_3(\kappa_i')[1-\kappa_i'(0)]\right) + (1-q_0)\left(\sum_{i=1}^{5} p_4(\kappa_i')\kappa_i'(0)\right) = C_S. \tag{H30}$$

[Here $\kappa_i'(a)$ denotes the ath entry of vertex $\kappa_i' = (\kappa_i'(0), \kappa_i'(1))$, where $i \in \{1,2,3,4,5\}$, $a \in \{1,2,3\}$].

We now carry out the remaining eliminations as follows. We construct the polytope of vectors $([p_2(\kappa_i')]_{i=1}^5, [p_3(\kappa_i')]_{i=1}^5, [p_4(\kappa_i')]_{i=1}^5, p_F, C_S, p_-)$ in $\mathbb{R}^{18}$ subject to the above constraints. This is a 12-dimensional polytope in $\mathbb{R}^{18}$ with 45 vertices, denoted *robustawvpolyN*.

We project the vertices down to just the coordinates (p_F, C_S, p_-) and construct a polytope with these as an input to *Polyhedron()*. This results in the polytope named *redawvpolyN*, a 3-dimensional polytope in $\mathbb{R}^3$ with 10 vertices.

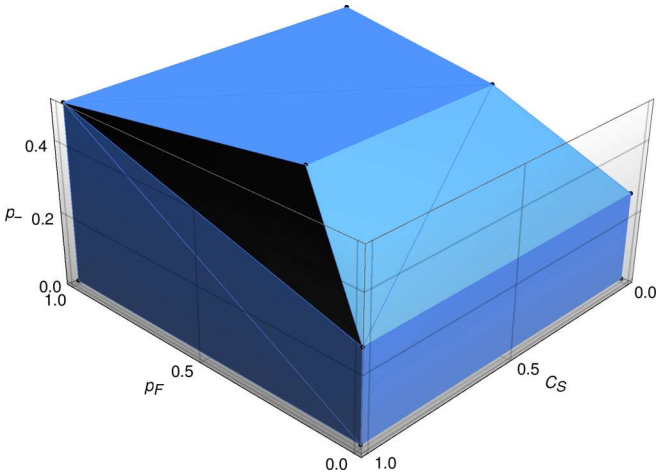

FIG. 5. The noncontextuality tradeoff between p_-, p_F and C_S for $p_d = 1/4$, $\tilde{p} = 1/2$, $q_0 = q_* = 1/2$. The facet corresponding to (37) is shown in black.

For a representative case of $(q_*, q_\perp, p_d, \tilde{p}) = (\frac{1}{2}, \frac{1}{2}, \frac{1}{4}, \frac{1}{2})$ the facets of this polytope include our noncontextuality inequality Eq. (C4): $p_F - 4C_S - 4p_- + 5 \geqslant 0$ or

$$p_- \leqslant \frac{p_F}{2} + \frac{1-p_F}{4} + \left(1-\frac{1}{2}\right)\frac{1-C_S}{1/2} = \tilde{p}p_F + p_d(1-p_F) + (1-\tilde{p})\frac{1-C_S}{p_*}. \tag{H31}$$

The overall tradeoff between (p_F, C_S, p_-) for this case is depicted in Fig. 5.

We also tried many other values of $(q_*, q_\perp, p_d, \tilde{p})$. Equation (C4) always appeared as a facet, except when $p_d > \tilde{p}$.

[1] Y. Aharonov, D. Z. Albert, and L. Vaidman, Phys. Rev. Lett. **60**, 1351 (1988).
[2] J. Dressel, M. Malik, F. M. Miatto, A. N. Jordan, and R. W. Boyd, Rev. Mod. Phys. **86**, 307 (2014).
[3] A. J. Leggett, Phys. Rev. Lett. **62**, 2325 (1989).
[4] Y. Aharonov and L. Vaidman, Phys. Rev. Lett. **62**, 2327 (1989).
[5] C. Ferrie and J. Combes, Phys. Rev. Lett. **113**, 120404 (2014).
[6] A. Brodutch, Phys. Rev. Lett. **114**, 118901 (2015).
[7] C. Ferrie and J. Combes, Phys. Rev. Lett. **114**, 118902 (2015).
[8] L. Vaidman, Philos. Trans. R. Soc. A **375**, 20160395 (2017).
[9] M. F. Pusey, Phys. Rev. Lett. **113**, 200401 (2014).
[10] R. W. Spekkens, Phys. Rev. A **71**, 052108 (2005).
[11] F. Piacentini, A. Avella, M. P. Levi, R. Lussana, F. Villa, A. Tosi, F. Zappa, M. Gramegna, G. Brida, I. P. Degiovanni, and M. Genovese, Phys. Rev. Lett. **116**, 180401 (2016).
[12] S. D. Bartlett, T. Rudolph, and R. W. Spekkens, Phys. Rev. A **86**, 012103 (2012).
[13] A. Karanjai, E. G. Cavalcanti, S. D. Bartlett, and T. Rudolph, New J. Phys. **17**, 073015 (2015).
[14] D. Schmid, R. W. Spekkens, and E. Wolfe, Phys. Rev. A **97**, 062103 (2018).
[15] R. Jozsa, Phys. Rev. A **76**, 044103 (2007).
[16] J. G. Kirkwood, Phys. Rev. **44**, 31 (1933).
[17] P. A. M. Dirac, Rev. Mod. Phys. **17**, 195 (1945).
[18] H. Margenau and R. N. Hill, Prog. Theor. Phys. **26**, 722 (1961).
[19] J. Dressel, Phys. Rev. A **91**, 032116 (2015).
[20] M. J. W. Hall, Phys. Rev. A **69**, 052113 (2004).
[21] M. Lostaglio, Phys. Rev. Lett. **120**, 040602 (2018).
[22] A. E. Allahverdyan, Phys. Rev. E **90**, 032137 (2014).
[23] S. Wu and K. Mølmer, Phys. Lett. A **374**, 34 (2009).
[24] D. Lu, A. Brodutch, J. Li, H. Li, and R. Laflamme, New J. Phys. **16**, 053015 (2014).
[25] R. Kunjwal and R. W. Spekkens, Phys. Rev. Lett. **115**, 110403 (2015).
[26] M. D. Mazurek, M. F. Pusey, R. Kunjwal, K. J. Resch, and R. W. Spekkens, Nat. Commun. **7**, 11780 (2016).
[27] M. F. Pusey, L. Del Rio, and B. Meyer, arXiv:1904.08699.
[28] N. Harrigan and R. W. Spekkens, Found. Phys. **40**, 125 (2010).
[29] C. Villani, *Topics in Optimal Transportation*, Graduate Studies in Mathematics Vol. 58 (American Mathematical Society, Providence, RI, 2003), p. 44.
[30] SageMath code, available at https://www.mattpusey.uk/weak.