# Conditions tighter than noncommutation needed for nonclassicality

David R. M. Arvidsson-Shukur,[1, 2, 3] Jacob Chevalier Drori,[4] and Nicole Yunger Halpern[5, 6, 3, 7, 8, 9]

[1] *Hitachi Cambridge Laboratory, J. J. Thomson Avenue, CB3 0HE, Cambridge, United Kingdom*
[2] *Cavendish Laboratory, Department of Physics, University of Cambridge, Cambridge CB3 0HE, United Kingdom*
[3] *Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[4] *DAMTP, Centre for Mathematical Sciences, University of Cambridge, Cambridge CB3 0WA, United Kingdom*
[5] *ITAMP, Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA*
[6] *Department of Physics, Harvard University, Cambridge, MA 02138, USA*
[7] *Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[8] *Joint Center for Quantum Information and Computer Science, NIST and University of Maryland, College Park, MD 20742, USA*
[9] *Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742, USA*

(Dated: January 31, 2021)

Kirkwood discovered in 1933, and Dirac discovered in 1945, a representation of quantum states that has undergone a renaissance recently. The Kirkwood-Dirac (KD) distribution has been employed to study nonclassicality across quantum physics, from metrology to chaos to the foundations of quantum theory. The KD distribution is a quasiprobability distribution, a quantum generalization of a probability distribution, which can behave nonclassically by having negative or nonreal elements. Negative KD elements signify quantum information scrambling and potential metrological quantum advantages. Nonreal elements encode measurement disturbance and thermodynamic nonclassicality. KD distributions' nonclassicality has been believed to follow necessarily from noncommutation of operators. We show that noncommutation does not suffice. We prove sufficient conditions for the KD distribution to be nonclassical (equivalently, necessary conditions for it to be classical). We also quantify the KD nonclassicality achievable under various conditions. This work resolves long-standing questions about nonclassicality and may be used to engineer quantum advantages.

*Introduction.*—Heisenberg's uncertainty principle [1–3] and Bohr's complementarity principle [4] power much of the strangeness in quantum mechanics. The principles codify the *incompatibility* of simultaneous measurements of certain observables. Despite incompatibility's essentiality in quantum physics, how the corresponding nonclassicality is best quantified remains unknown [5]. Guided by practicality, we use Kirkwood and Dirac's quasiprobability formalism of quantum mechanics [6, 7], reviewed below. We prove how operator incompatibility underlies, but does not guarantee, negative and nonreal quasiprobabilities, which signal nonclassical physics under certain circumstances. We then quantify and bound the distribution's nonclassicality.

In classical mechanics, a joint probability-density function $\mathcal{P}(\boldsymbol{x}, \boldsymbol{p})$ describes a system's position $\boldsymbol{x}$ and momentum $\boldsymbol{p}$. In quantum mechanics, observables do not necessarily commute. Representing a state with a joint probability function over observables' eigenvalues is generally impossible [8–13].

By forfeiting one of Kolmogorov's axioms of joint probability functions [14], one can represent quantum mechanics with a probability-like framework. A quantum state can be represented by a quasiprobability function over incompatible observables' eigenvalues. A quasiprobability behaves like a probability but can assume negative and/or nonreal values. Many types of quasiprobability distributions exist. The best-known is the Wigner function, a function of position and momentum [8, 15, 16]. The Wigner function (and the related Sudarshan-Glauber P and Husimi Q representations [17–19]) are used extensively in quantum optics [20], where $\boldsymbol{x}$ and $\boldsymbol{p}$ are swapped for the electric field's the real and imaginary components. However, in experiments that lack clear analogs of $\boldsymbol{x}$ and $\boldsymbol{p}$, the Wigner function is less suitable. Furthermore, Wigner-function negativity is neither necessary nor sufficient for nonclassical phenomena: The Einstein-Podolsky-Rosen state [21] has a positive Wigner function [22], and states expressibly classically in the particle-number basis can have negative Wigner representations [23].

The Kirkwood-Dirac[1] (KD) quasiprobability distribution is a relative of the Wigner function. Kirkwood [6] and Dirac [7] independently developed the KD distribution to facilitate the application of probability theory to quantum mechanics. Compared to the Wigner function, the KD distribution possesses an additional freedom: It can assume nonreal values. Moreover, the KD distribution is straightforwardly defined for discrete systems—even qubits.

The KD distribution has recently illuminated several areas of quantum mechanics. In weak-value amplification [28–31], negative KD quasiprobabilities allow pre- and postselected averages of observables, *weak values*, to lie outside the obervables' eigenspectra, improving signal-to-noise ratios [32–39]. Nonreal KD quasiprobabilities can endow weak values with imaginary components,

[1] The Kirkwood-Dirac distribution has been called by several names. Its real part is often called the Terletsky-Margenau-Hill distribution [24–27].

which encode a measurement's disturbance of a quantum state [40–43]. Measuring a KD distribution allows for the tomographic reconstruction of a quantum state [44–48]. In quantum chaos, quantum-information scrambling (the spreading of a local perturbation via many-body entanglement) is quantified with an out-of-time-ordered correlator [49, 50]. This correlator drops to classically forbidden values when underlying KD quasiprobabilities assume negative or nonreal values [38, 51–53, 68]. In quantum metrology, postselection can increase the average amount of information obtained about an unknown parameter per end-of-trial measurement [54–56]. If the postselection is designed such that a conditional KD distribution contains negative elements, the information-per-final-measurement rate can be nonclassically large. KD distributions have been used in quantum thermodynamics [57, 58, 68]; nonreal KD quasiprobabilities enable an engine to be unexplainable by any classical (noncontextual) theory [58]. Finally, the KD distribution has applications to the foundations of quantum mechanics [12, 41, 59–67]. For example, a KD distribution is related to histories' weights in the consistent-histories interpretation of quantum mechanics [12, 59, 60].

Despite the KD distribution's versatility, many of its properties have not been detailed. A common misconception is that noncommutation guarantees negative or nonreal KD quasiprobabilities. Furthermore, no bounds are known on how much nonclassicality a KD distribution can have. An improved understanding of the KD distribution's properties can facilitate the design of diverse experiments that harness the distribution's nonclassicality for quantum advantage.

In this Article, we prove sufficient conditions for the KD distribution to have nonclassically negative and/or nonreal values (Thm. 1) or, equivalently, necessary conditions for the KD distribution to be classical. We identify cases in which the KD distribution is classical despite pairwise noncommutation between the quantum state and the observables in the distribution's definition. Our results extend to scenarios where the KD distribution is coarse-grained to account for degeneracies in experiments. Reference [52] introduced a measure for the KD distribution's nonclassicality. We complement this measure with new ones, suited to more-diverse operational tasks. We also upper-bound these nonclassicality measures (Thm. 2). Conditioning the KD distribution, à la Bayes' theorem, allows KD nonclassicality to exceed the bounds, amplifying quantum advantages in certain experiments. Finally, we quantify how decoherence reduces KD distributions' nonclassicalities.

*Kirkwood-Dirac distribution.*—We assume that all operators operate on a Hilbert space with finite dimension $d$. Consider two orthonormal bases, $\{|a_i\rangle\}$ and $\{|f_i\rangle\}$. Throughout this article, we regard these bases as eigenbases of observables $\hat{A} = \sum_i a_i |a_i\rangle\langle a_i|$ and $\hat{F} = \sum_i f_i |f_i\rangle\langle f_i|$. In terms of these bases, a state $\hat{\rho}$ can be represented by the KD distribution

$$\left\{q^{\hat{\rho}}_{i,j}\right\} \equiv \{\langle f_j|a_i\rangle\langle a_i|\hat{\rho}|f_j\rangle\} = \left\{\mathrm{Tr}(\hat{\Pi}^f_j\hat{\Pi}^a_i\hat{\rho})\right\}, \qquad (1)$$

where $\hat{\Pi}^a_i \equiv |a_i\rangle\langle a_i|$, etc. The distribution can be used to calculate expectation values and measurement-outcome probabilities. $\{q^{\hat{\rho}}_{i,j}\}$ satisfies some of Kolmogorov's axioms for joint probability distributions [14]:

$$\sum_{i,j} q^{\hat{\rho}}_{i,j} = 1, \ \sum_j q^{\hat{\rho}}_{i,j} = p(a_i|\hat{\rho}), \text{ and } \sum_i q^{\hat{\rho}}_{i,j} = p(f_j|\hat{\rho}),$$

where $p(a_i|\hat{\rho})$ and $p(f_j|\hat{\rho})$ denote conditional probabilities. $q^{\hat{\rho}}_{i,j}$ can be nonclassical by assuming negative or nonreal values. Nonclassical values are not directly observable but cause effects inferable from sequential measurements [38]. If $\{|a_i\rangle\} = \{|f_j\rangle\}$, the KD distribution reduces to a classical probability distribution: $\{q^{\hat{\rho}}_{i,j}\} = \{\langle f_j|a_i\rangle\langle a_i|\hat{\rho}|f_j\rangle\,\delta_{f_j,a_i}\} = \{\mathrm{Tr}(\hat{\Pi}^a_i\hat{\rho})\delta_{f_j,a_i}\}$. In classical physics, all observables commute, and every KD distribution equals a probability distribution.

Certain physical processes [38, 51–53, 56, 68] motivate the extension of the KD distribution from 2 to $k$ bases, e.g., eigenbases of $k$ observables $\hat{A}^{(1)}, \ldots, \hat{A}^{(k)}$. The extended KD distribution is

$$\left\{q^{\hat{\rho}}_{i_1,\ldots,i_k}\right\} \equiv \left\{\mathrm{Tr}\left(\hat{\Pi}^{a^{(k)}}_{i_k}\ldots\hat{\Pi}^{a^{(1)}}_{i_1}\hat{\rho}\right)\right\}. \qquad (2)$$

A KD distribution's elements serve as the coefficients in an operator expansion of $\hat{\rho}$:

$$\hat{\rho} = \sum_{i_1,\ldots,i_k}\frac{|a^{(1)}_{i_1}\rangle\langle a^{(k)}_{i_k}|}{\langle a^{(k)}_{i_k}|a^{(1)}_{i_1}\rangle}q^{\hat{\rho}}_{i_1,\ldots,i_k} = \sum_{i,j}\frac{|a^{(1)}_i\rangle\langle a^{(k)}_j|}{\langle a^{(k)}_j|a^{(1)}_i\rangle}q^{\hat{\rho}}_{i,j}. \quad (3)$$

We define $q^{\hat{\rho}}_{i,j}/\langle a^{(k)}_j|a^{(1)}_i\rangle \equiv \langle a^{(1)}_i|\hat{\rho}|a^{(k)}_j\rangle$ if $\langle a^{(k)}_j|a^{(1)}_i\rangle = 0$.

We have shown how to represent a state in terms of eigenbases of Hermitian operators, including measured observables and time-evolution generators. In terms of this representation, physical quantities can be expressed. Assuming that KD distributions are real and non-negative, one can bound the values attainable in classical settings. This strategy has been applied to weak values[2] [38, 69], information scrambling [52, 53], and the Fisher information [56]. Nonclassicality in the KD distribution is a stricter condition than noncommutation, we show, as the former requires the latter but not *vice versa*.

*Requirement for nonclassical quasiprobabilities.*—If any two of $\hat{A}$, $\hat{F}$, and $\hat{\rho}$ commute, they share at least one eigenbasis. When $\hat{A}$ and $\hat{F}$ commute and a shared eigenbasis serves as the $\{|a_i\rangle\}$ and the $\{|f_j\rangle\}$ in Eq. (1),

[2] Observables' expectation values equal KD-weighted weak values [62].

the KD distribution equals a classical probability distribution. When $\hat{\rho}$ and $\hat{A}$ ($\hat{F}$) commute, it suffices for classicality that a shared eigenbasis serves as $\{|a_i\rangle\}$ ($\{|f_j\rangle\}$). If $[\hat{\rho}, \hat{A}], [\hat{\rho}, \hat{F}], [\hat{A}, \hat{F}] \neq 0$, the KD distribution *may* assume negative or nonreal values. However, noncommutation does not suffice for KD nonclassicality, as shown in Examples 1 and 2 in App. A. To find a sufficient condition for nonclassicality (equivalently, a necessary condition for classicality), we focus first on (i) pure states $\hat{\rho}$ and (ii) nondegenerate $\hat{A}$ and $\hat{F}$. We then address degenerate observables and mixed states.

Let us define four real numbers that reflect incompatibility properties of $\hat{\rho}$, $\hat{A}$, and $\hat{F}$. In the pure case, $\hat{\rho} = |\Psi\rangle\langle\Psi|$. Let $\mathcal{V}_A \equiv \{|a_i\rangle\}$ and $\mathcal{V}_F \equiv \{|f_j\rangle\}$ denote the eigenbases of the nondegenerate $\hat{A}$ and $\hat{F}$, respectively. These eigenbases are unique up to phases. Define as $N_A$ ($N_F$) the number of $\mathcal{V}_A$ ($\mathcal{V}_F$) vectors whose overlaps with $|\Psi\rangle$ are nonzero:

$$N_A \equiv ||\{|a_i\rangle \in \mathcal{V}_A : \langle a_i|\Psi\rangle \neq 0\}||, \text{ and} \tag{4}$$

$$N_F \equiv ||\{|f_j\rangle \in \mathcal{V}_F : \langle f_j|\Psi\rangle \neq 0\}||. \tag{5}$$

$||\cdot||$ denotes a set's cardinality. We denote by $n_\parallel$ ($\bar{n}_\parallel$) the number of $|a_i\rangle$ that are (i) parallel to vectors $|f_j\rangle$ and (ii) nonorthogonal (orthogonal) to $|\Psi\rangle$.

**Theorem 1** (Sufficient conditions for Kirkwood-Dirac nonclassicality)**.** *Suppose that $\hat{\rho}$ is pure and that $\hat{A}$ and $\hat{F}$ are nondegenerate. If $2N_A + 2N_F > 3d + n_\parallel - 3\bar{n}_\parallel$, then the Kirkwood-Dirac distribution contains negative or nonreal values.*

We prove the theorem by the contrapositive: Assuming a classical KD distribution, we deduce constraints on the unitary matrix with entries $\langle a_i|f_j\rangle$. These constraints imply a condition on $N_A$, $N_F$, $d$, $n_\parallel$, and $\bar{n}_\parallel$ that is necessary for classicality of the KD distribution. A violation of this condition suffices for KD nonclassicality. The full proof appears in App. B

Theorem 1 implies a simple condition sufficient for KD nonclassicality:

**Corollary 1.** *If the KD distribution lacks zero-valued quasiprobabilities, $\{q^{\hat{\rho}}_{i,j}\}$ is nonclassical.*

*Proof:* If all $q^{\hat{\rho}}_{i,j} \neq 0$, then $|a_i\rangle \not\parallel |f_j\rangle$,[3] and $\langle a_i|\Psi\rangle, \langle f_j|\Psi\rangle \neq 0$, for all $i, j$. So $n_\parallel = \bar{n}_\parallel = 0$, and $N_A = N_F = d$, satisfying the nonclassicality condition of Thm. 1. □

Three more extensions of Thm. 1 merit mention. First, if $\hat{A}$ and $\hat{F}$ are degenerate, one can construct KD distributions by coarse-graining over the degeneracies. These coarse-grained distributions can signal nonclassical physics in quantum chaos [38, 51–53] and metrology [56]. In App. D, we prove sufficient conditions for these distributions to be nonclassical.

Second, every KD distribution $\{q^{\hat{\rho}}_{i_1,i_k}\}$ follows from marginalizing an extended distribution $\{q^{\hat{\rho}}_{i_1,\ldots,i_k}\}$ [Eq. (2)] over the indices $i_2, \ldots, i_{k-1}$ [38, 51–53, 56, 68]. If any marginalized $\{q^{\hat{\rho}}_{i_\alpha,i_\beta}\}$ satisfies the nonclassicality condition in Thm. 1, every fine-graining $\{q^{\hat{\rho}}_{i_1,\ldots,i_k}\}$ is nonclassical.

Third, we prove further properties of the real and imaginary components of $q^{\hat{\rho}}_{i,j}$ in App. C. These properties can be used, e.g., to tailor states $\hat{\rho}$ to achieve nonclassical results in experiments that involve observables $\hat{A}$ and $\hat{F}$. A similar strategy is being applied in a photonic experiment to observe how KD negativity benefits parameter estimation [70].

*Nonclassicality measures.*—How much nonclassicality can a KD distribution have? We review an existing nonclassicality measure, define measures suited to more operational tasks, and upper-bound the measures.

Every KD distribution's elements sum to unity. Negative and nonreal entries are nonclassical. González Alonso *et al.* thus quantified [52] KD distributions' nonclassicality, in the context of scrambling, with

$$\mathcal{N}\left(\{q^{\hat{\rho}}_{i_1,\ldots,i_k}\}\right) \equiv -1 + \sum_{i_1,\ldots,i_k} |q^{\hat{\rho}}_{i_1,\ldots,i_k}|. \tag{6}$$

$\mathcal{N}(\{q^{\hat{\rho}}_{i_1,\ldots,i_k}\}) = 0$ when $\{q^{\hat{\rho}}_{i_1,\ldots,i_k}\}$ is real and nonnegative. We upper-bound the measure generally in terms of the Hilbert-space dimensionality, $d$.

**Theorem 2** (Maximum Kirkwood-Dirac nonclassicality)**.** *The maximum nonclassicality $\mathcal{N}(\{q^{\hat{\rho}}_{i_1,\ldots,i_k}\})$ of any Kirkwood-Dirac distribution $\{q^{\hat{\rho}}_{i_1,\ldots,i_k}\}$ is*

$$\max_{\hat{\rho},\hat{A}^{(1)},\ldots,\hat{A}^{(k)}} \left\{\mathcal{N}\left(\{q^{\hat{\rho}}_{i_1,\ldots,i_k}\}\right)\right\} = d^{(k-1)/2} - 1. \tag{7}$$

*The maximum is achieved if and only if two conditions are met simultaneously: (i) The operators $\hat{A}^{(i)}$ and $\hat{A}^{(i+1)}$ have mutually unbiased eigenbases*[4] *(MUBs) for each $i = 1, \ldots, k-1$. (ii) $\hat{\rho} = |\Psi\rangle\langle\Psi|$, where $|\Psi\rangle$ has equal overlaps with all the eigenvectors of $\hat{A}^{(1)}$ and $\hat{A}^{(k)}$.*

The proof of Thm. 2 appears in App. E.

At least one triplet of MUBs exists for every $d \geq 2$ [71]. We can therefore construct a $\left\{q^{|\Psi\rangle\langle\Psi|}_{i_1,\ldots,i_k}\right\}$ that maximizes $\mathcal{N}$: Let $|\Psi\rangle$ be an element of the triplet's first MUB. Let $|a^{(k)}_{i_k}\rangle$ be the $i_k^{\text{th}}$ element of the second (third) MUB if $k$ is even (odd).

---

[3] Since $\mathcal{V}_A$ and $\mathcal{V}_F$ are orthonormal sets, if some $|a_i\rangle \parallel |f_j\rangle$, then some other $|a_{i'}\rangle \perp |f_j\rangle$. By Eq. (1), $q^{\hat{\rho}}_{i',j} = 0$.

[4] Bases $\mathcal{A} \equiv \{|\alpha_j\rangle\}$ and $\mathcal{B} \equiv \{|\beta_k\rangle\}$ are *mutually unbiased* if preparing any $\mathcal{A}$ element and measuring $\mathcal{B}$ yields a totally unpredictable outcome: $|\langle\alpha_j|\beta_k\rangle| = 1/\sqrt{d}$ for all $j, k$.

The measure (6) is useful in the context of chaos, where negative and nonreal KD quasiprobabilities signal scrambling [53]. But negative and nonreal values do not always enjoy equal footing: Only negative KD quasiprobabilities enable a metrologist to garner a nonclassically high Fisher information [56]. In contrast, nonreal KD quasiprobabilities lie behind weak values' imaginary components, which encode measurement disturbance [32, 42]. We therefore quantify the aggregated negativity and nonreality, respectively:

$$\mathcal{N}^{\Re^-}\left(\{q^{\hat{\rho}}_{i_1,\ldots,i_k}\}\right) := -1 + \sum_{i_1,\ldots,i_k} \left|\Re(q^{\hat{\rho}}_{i_1,\ldots,i_k})\right|, \text{ and} \quad (8)$$

$$\mathcal{N}^{\Im}\left(\{q^{\hat{\rho}}_{i_1,\ldots,i_k}\}\right) := \sum_{i_1,\ldots,i_k} \left|\Im(q^{\hat{\rho}}_{i_1,\ldots,i_k})\right|. \quad (9)$$

$\mathcal{N}^{\Re^-} \leq \mathcal{N}$ by definition, and $0 \leq \mathcal{N}^{\Im} < \mathcal{N} + 1$. If all the nonclassical $q^{\hat{\rho}}_{i_1,\ldots,i_k}$ are real negative numbers, $\mathcal{N}^{\Re^-} = \mathcal{N}$. Given the importance of $\mathcal{N}^{\Re^-}$ to quantum metrology and weak-value amplification, a crucial question is: When can $\mathcal{N}^{\Re^-} = \max\{\mathcal{N}\}$? A complete answer requires further advances in the field of MUBs. Nevertheless, for every $d$ in which a triplet of real MUBs exists,[5] $\max\{\mathcal{N}^{\Re^-}\} = \max\{\mathcal{N}\}$. The number of real MUBs in a space of a general dimensionality $d$ is unknown. The smallest space with a triplet of real MUBs has $d = 4$ [72]. We construct an example in which $d = 4$ and $\max\{\mathcal{N}^{\Re^-}\} = \max\{\mathcal{N}\}$ in Ex. 3 of App. A. In $d = 2$, the Pauli bases form a triplet of MUBs. When $k = 2$ and the Pauli bases are used to maximize $\mathcal{N}$, all nonclassicality manifests as nonreal quasiprobabilities without negative real components (App. A, Ex. 4).

*Amplifying nonclassicality via postselection.*—As aforementioned, negative KD quasiprobabilities underlie quantum advantages in weak-value amplification and postselected quantum metrology. The reason is, the protocols involve postselection. Classical postselection, or conditioning, obeys Bayes' theorem, $p(a|b) = p(b|a)p(a)/p(b)$. The KD distribution satisfies an analog of Bayes' theorem [26, 38, 44]: Suppose that a state represented by $\{q^{\hat{\rho}}_{i_1,\ldots,i_k}\}$ undergoes a measurement $\{\hat{F}_k, \hat{1}-\hat{F}_k\}$, where $\hat{F}_k \equiv \sum_{i_k : |f_{i_k}\rangle \in \mathcal{F}_k} |f_{i_k}\rangle\langle f_{i_k}|$ for some set $\mathcal{F}_k$. Conditioned on the outcome's corresponding to $\hat{F}_k$, the KD quasiprobabilities are

$$\frac{\sum_{i_k : |f_{i_k}\rangle \in \mathcal{F}_k} q^{\hat{\rho}}_{i_1,\ldots,i_k}}{p(F_k|\hat{\rho})}, \quad \text{where} \quad (10)$$

$$p(F_k|\hat{\rho}) \equiv \sum_{\substack{i_1,\ldots,i_{k-1},\\ i_k : |f_{i_k}\rangle \in \mathcal{F}_k}} q^{\hat{\rho}}_{i_1,\ldots,i_k} = \mathrm{Tr}(\hat{F}_k \hat{\rho}). \quad (11)$$

The form of $q^{\hat{\rho}}_{i_1,\ldots,i_k}$ [Eq. (1)] implies that, for every unconditioned KD distribution, $0 \leq |q^{\hat{\rho}}_{i_1,\ldots,i_k,j}| \leq 1$. If $\{q^{\hat{\rho}}_{i_1,\ldots,i_k}\}$ lacks nonclassical values, also the conditional KD quasiprobabilities (10) lie between 0 and 1. However, if $q^{\hat{\rho}}_{i_1,\ldots,i_k}$ contains negative values, the numerator in Eq. (10) can have a greater magnitude than the denominator. The conditional quasiprobability can be made arbitrarily large [56]. So can, consequently, the corresponding $\mathcal{N}$, $\mathcal{N}^{\Re^-}$, and $\mathcal{N}^{\Im}$. This KD nonclassicality can lead to metrological capabilities infinitely greater than those achievable classically [sometimes at a cost of low postselection probabilities $p(F_k|\hat{\rho})$] [28, 29, 56].

*Mixed states.*—We have focused on pure-state KD distributions, but every experiment involves decoherence. How does decoherence affect KD nonclassicality? Let $\hat{\rho} = \sum_n p_n \hat{\rho}_n$, where $\hat{\rho}_n \equiv |\Psi_n\rangle\langle\Psi_n|$ and $p_n$ denotes a probability. $\hat{\rho}$ can be represented by the KD distribution$\{q^{\hat{\rho}}_{i,j}\} = \{\sum_n p_n q^{\hat{\rho}_n}_{i,j}\}$. By convexity, the nonclassical $q^{\hat{\rho}}_{i,j}$ have magnitudes no greater than the magnitudes of the nonclassical components of the most nonclassical $\{q^{\hat{\rho}_n}_{i,j}\}$: Mixing dilutes the nonclassicality. For example, the KD distributions for the pure states $\hat{\rho}_+ = |+\rangle\langle+|$ and $\hat{\rho}_- = |-\rangle\langle-|$ with respect to the bases $\{|a\rangle\} = \{|0\rangle, |1\rangle\}$ and $\{|f\rangle\} = \{\cos(\pi/3)|0\rangle + \sin(\pi/3)|1\rangle, -\sin(\pi/3)|0\rangle + \cos(\pi/3)|1\rangle\}$ are nonclassical. But the distribution for $\hat{\rho} = \frac{2}{3}\hat{\rho}_+ + \frac{1}{3}\hat{\rho}_-$ is classical.[6] Decoherence obscures the incompatible eigenbases' nonclassicality.

In another example, consider depolarizing a pure state $\hat{\rho}_0$: $\hat{\rho}' \equiv p\hat{\rho}_0 + (1-p)\hat{1}/d$. The KD distribution of $\hat{\rho}'$ has elements

$$q^{\hat{\rho}'}_{i,j} = p\, q^{\hat{\rho}_0}_{i,j} + \frac{1-p}{d} |\langle f_j|a_i\rangle|^2. \quad (12)$$

If $p$ is small enough (e.g., if $p = 0$), the depolarizing channel eliminates the KD distribution's negative components. By the triangle inequality, $\mathcal{N}(\{q^{\hat{\rho}'}_{i,j}\}) \leq p\mathcal{N}(\{q^{\hat{\rho}_0}_{i,j}\})$, and $\mathcal{N}^{\Re^-}(\{q^{\hat{\rho}'}_{i,j}\}) \leq p\mathcal{N}^{\Re^-}(\{q^{\hat{\rho}_0}_{i,j}\})$. Each imaginary component is reduced by a factor of $p$: $\mathcal{N}^{\Im}(\{q^{\hat{\rho}'}_{i,j}\}) = p\mathcal{N}^{\Im}(\{q^{\hat{\rho}_0}_{i,j}\})$. $\mathcal{N}^{\Im}(\{q^{\hat{\rho}'}_{i,j}\})$ can resist decoherence more than $\mathcal{N}^{\Re^-}(\{q^{\hat{\rho}'}_{i,j}\})$: Only when the state decoheres fully ($p = 0$) do all the imaginary components disappear. The negative components disappear when the decoherence surpasses a finite threshold.

*Discussion.*—Benefits of using the KD distribution include the ability to prove classical bounds on physical quantities by assuming real, non-negative distributions. The key to applying the KD distribution fruitfully is to construct the distribution operationally. The bases and their ordering should reflect properties of the experiment (e.g., [38, 53, 56, 58, 70]). Similarly, experimental context

[5] For our purposes, a real MUB is an MUB whose vectors can be expressed, relative to a fixed basis, as columns of real numbers. Appendix F reconciles this definition with the conventional definition.

[6] $|+\rangle$ ($|-\rangle$) and $|0\rangle$ ($|1\rangle$) denote the $+1$ ($-1$) eigenvectors of the Pauli-$x$ and Pauli-$z$ operators, respectively.

dictates when extending the KD distribution facilitates analyses [38, 51–53, 56, 68].

Our work provides a methodology for calculating whether an input state and subsequent operations may generate nonclassical physics in a range of experiments. Furthermore, our work provides a mathematical toolkit for constructing quantum-enhanced experiments. We have shown that noncommutation does not suffice for achieving nonclassical KD distributions and associated quantum advantages. Instead, KD negativity and nonreality emerge as sharper nonclassicality criteria than noncommutation for diverse tasks.

*Acknowledgements.*—The authors would like to thank Crispin Barnes, Stephan de Bièvre, Nicolas Delfosse, Giacomo De Palma, and Justin Dressel for useful discussions. D.R.M.A.-S. was supported by the EPSRC, Lars Hierta's Memorial Foundation, and Girton College. N.Y.H. was supported by an NSF grant for the Institute for Theoretical Atomic, Molecular, and Optical Physics at Harvard University and the Smithsonian Astrophysical Observatory and by the MIT CTP administratively.

## SUPPLEMENTARY MATERIAL

### Appendix A: Example KD distributions

**Example 1** (Classical KD distribution for pairwise-noncommuting $\hat{A}$, $\hat{F}$, and pure $\hat{\rho}$)**.** *Consider a two-qubit system. As before, $|+\rangle$ ($|-\rangle$) and $|0\rangle$ ($|1\rangle$) are the $+1$ ($-1$) eigenvectors of the Pauli-x and Pauli-z operators, respectively. We choose $\hat{A}$ and $\hat{F}$ such that $\{|a_i\rangle\} = \{|0\rangle|0\rangle, |0\rangle|1\rangle, |1\rangle|0\rangle, |1\rangle|1\rangle\}$ and $\{|f_j\rangle\} = \{|0\rangle|+\rangle, |0\rangle|-\rangle, |1\rangle|0\rangle, |1\rangle|1\rangle\}$. For example, if each observable has the eigenvalues $-2$, $-1$, $1$, and $2$,*

$$\hat{A} \to \begin{pmatrix} -2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}, \text{ and } \hat{F} \to \begin{pmatrix} -\frac{3}{2} & -\frac{1}{2} & 0 & 0 \\ -\frac{1}{2} & -\frac{3}{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}. \tag{A1}$$

*We set $\hat{\rho} = |\Psi\rangle\langle\Psi|$, where $|\Psi\rangle = |1\rangle|+\rangle$:*

$$\hat{\rho} \to \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}. \tag{A2}$$

*$\hat{A}$, $\hat{F}$ and $\hat{\rho}$ fail to commute pairwise: $[\hat{A}, \hat{F}]$, $[\hat{\rho}, \hat{A}]$, $[\hat{\rho}, \hat{F}] \neq 0$. However, the KD distribution (Table I) is real and non-negative.*

TABLE I. The KD distribution of Ex. 1.

| $\lvert f_j\rangle$ \ $\lvert a_i\rangle$ | $\lvert 0\rangle\lvert 0\rangle$ | $\lvert 0\rangle\lvert 1\rangle$ | $\lvert 1\rangle\lvert 0\rangle$ | $\lvert 1\rangle\lvert 1\rangle$ |
|---|---|---|---|---|
| $\lvert 0\rangle\lvert +\rangle$ | 0 | 0 | 0 | 0 |
| $\lvert 0\rangle\lvert -\rangle$ | 0 | 0 | 0 | 0 |
| $\lvert 1\rangle\lvert 0\rangle$ | 0 | 0 | $\frac{1}{2}$ | 0 |
| $\lvert 1\rangle\lvert 1\rangle$ | 0 | 0 | 0 | $\frac{1}{2}$ |

*Since this KD distribution is classical, Thm. 1 implies that $2N_A + 2N_F \leq 3d + n_\parallel - 3\bar{n}_\parallel$. Indeed, $N_A = N_F = 2$, $d = 4$, $n_\parallel = 2$, and $\bar{n}_\parallel = 0$; so the inequality reads $8 \leq 14$.*

**Example 2** (Classical KD distribution that saturates Ineq. (B2))**.** *Consider a 4-dimensional Hilbert space with an orthonormal basis $\{|0\rangle, |1\rangle, |2\rangle, |3\rangle\}$. Suppose that $\hat{A}$ and $\hat{F}$ have eigenbases $\{|a_i\rangle\} = \{|0\rangle, |1\rangle, |2\rangle, |3\rangle\}$ and $\{|f_j\rangle\} = \{\frac{|0\rangle+|1\rangle}{\sqrt{2}}, \frac{|0\rangle-|1\rangle}{\sqrt{2}}, \frac{|2\rangle+|3\rangle}{\sqrt{2}}, \frac{|2\rangle-|3\rangle}{\sqrt{2}}\}$. Let $\hat{\rho} = |\Psi\rangle\langle\Psi|$, where $|\Psi\rangle = \frac{|0\rangle+|1\rangle+|2\rangle+|3\rangle}{2}$. The KD distribution, presented in Table II, is real and non-negative.*
*In this example, $N_A = 4$, $N_F = 2$, $d = 4$, and $n_\parallel = \bar{n}_\parallel = 0$. Hence, $2N_A + 2N_F = 12 = 3d + n_\parallel - 3\bar{n}_\parallel$: The classical inequality $2N_A + 2N_F \leq 3d + n_\parallel - 3\bar{n}_\parallel$ obtained from Thm. 1 is saturated.*

TABLE II. The KD distribution of Ex. 2.

| $\lvert f_j\rangle$ \ $\lvert a_i\rangle$ | $\lvert 0\rangle$ | $\lvert 1\rangle$ | $\lvert 2\rangle$ | $\lvert 3\rangle$ |
|---|---|---|---|---|
| $\frac{\lvert 0\rangle+\lvert 1\rangle}{\sqrt{2}}$ | $\frac{1}{4}$ | $\frac{1}{4}$ | 0 | 0 |
| $\frac{\lvert 0\rangle-\lvert 1\rangle}{\sqrt{2}}$ | 0 | 0 | 0 | 0 |
| $\frac{\lvert 2\rangle+\lvert 3\rangle}{\sqrt{2}}$ | 0 | 0 | $\frac{1}{4}$ | $\frac{1}{4}$ |
| $\frac{\lvert 2\rangle-\lvert 3\rangle}{\sqrt{2}}$ | 0 | 0 | 0 | 0 |

**Example 3** (Real nonclassical KD distribution that achieves the maximum in Thm. 2)**.** *Suppose that* $\hat{A}$ *and* $\hat{F}$ *act on a two-qubit Hilbert space and have eigenbases* $\{\lvert a_i\rangle\} = \{\lvert 0\rangle\lvert 0\rangle, \lvert 0\rangle\lvert 1\rangle, \lvert 1\rangle\lvert 0\rangle, \lvert 1\rangle\lvert 1\rangle\}$ *and* $\{\lvert f_j\rangle\} = \{\lvert +\rangle\lvert +\rangle, \lvert -\rangle\lvert +\rangle, \lvert +\rangle\lvert -\rangle, \lvert -\rangle\lvert -\rangle\}$. $\{\lvert a_i\rangle\}$ *and* $\{\lvert f_j\rangle\}$ *form a pair of MUBs. Let* $\hat{\rho} = \lvert\Psi\rangle\langle\Psi\rvert$*, where* $\lvert\Psi\rangle = (\lvert 0\rangle\lvert 0\rangle + \lvert 0\rangle\lvert 1\rangle + \lvert 1\rangle\lvert 0\rangle - \lvert 1\rangle\lvert 1\rangle)/2$*. The overlaps* $\lvert\langle\Psi|a_i\rangle\rvert = \lvert\langle\Psi|f_j\rangle\rvert = \lvert\langle a_i|f_j\rangle\rvert = \frac{1}{2}$ *for all* $i, j$*. The resulting KD distribution is given in Table III.*

TABLE III. The KD distribution of Ex. 3.

| $\lvert f_j\rangle$ \ $\lvert a_i\rangle$ | $\lvert 0\rangle\lvert 0\rangle$ | $\lvert 0\rangle\lvert 1\rangle$ | $\lvert 1\rangle\lvert 0\rangle$ | $\lvert 1\rangle\lvert 1\rangle$ |
|---|---|---|---|---|
| $\lvert +\rangle\lvert +\rangle$ | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{1}{8}$ | $-\frac{1}{8}$ |
| $\lvert -\rangle\lvert +\rangle$ | $\frac{1}{8}$ | $\frac{1}{8}$ | $-\frac{1}{8}$ | $\frac{1}{8}$ |
| $\lvert +\rangle\lvert -\rangle$ | $\frac{1}{8}$ | $-\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{1}{8}$ |
| $\lvert -\rangle\lvert -\rangle$ | $-\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{1}{8}$ |

*This KD distribution is nonclassical. Furthermore,* $\mathcal{N}\left(\left\{q_{i,j}^{\lvert\Psi\rangle\langle\Psi\rvert}\right\}\right) = 1$ *saturates the inequality* $\mathcal{N}\left(\left\{q_{i_1,\ldots,i_k}^{\lvert\Psi\rangle\langle\Psi\rvert}\right\}\right) \leq d^{(k-1)/2} - 1$ *in Thm. 2, for* $k = 2$ *and* $d = 4$*. As the KD distribution is real, it saturates also* $\mathcal{N}^{\Re^-} \leq \mathcal{N}$*.*

**Example 4** (Nonclassical KD distribution for Pauli operators)**.** *Let* $\hat{A} = \hat{\sigma}_z$*,* $\hat{F} = \hat{\sigma}_x$*, and* $\hat{\rho} = \lvert\Psi\rangle\langle\Psi\rvert$*, where* $(\lvert 0\rangle + i\lvert 1\rangle)/\sqrt{2}$ *(the +1 eigenstate of* $\hat{\sigma}_y$*). The resulting KD distribution is given in Table IV.*

TABLE IV. The KD distribution of Ex. 4.

| $\lvert f_j\rangle$ \ $\lvert a_i\rangle$ | $\lvert 0\rangle$ | $\lvert 1\rangle$ |
|---|---|---|
| $\lvert +\rangle$ | $(1-i)/4$ | $(1+i)/4$ |
| $\lvert -\rangle$ | $(1+i)/4$ | $(1-i)/4$ |

*This KD distribution is nonclassical. Furthermore,* $\mathcal{N}\left(\left\{q_{i,j}^{\lvert\Psi\rangle\langle\Psi\rvert}\right\}\right) = \sqrt{2} - 1$ *saturates the inequality* $\mathcal{N}\left(\left\{q_{i,j}^{\lvert\Psi\rangle\langle\Psi\rvert}\right\}\right) \leq d^{(k-1)/2} - 1$ *in Thm. 2, for* $k = 2$ *and* $d = 2$*. The KD distribution is non-negative, so* $\mathcal{N}^{\Re^-}\left(\left\{q_{i,j}^{\lvert\Psi\rangle\langle\Psi\rvert}\right\}\right) = 0$*. All the nonclassicality lies in the imaginary components of* $\left\{q_{i,j}^{\lvert\Psi\rangle\langle\Psi\rvert}\right\}$*:* $\mathcal{N}^{\Im}\left(\left\{q_{i,j}^{\lvert\Psi\rangle\langle\Psi\rvert}\right\}\right) = 1$*. The results below Table IV hold for every version of the* $k = 2$ *KD distribution, where* $\{\lvert a_i\rangle\}$ *is one Pauli basis,* $\{\lvert f_j\rangle\}$ *is another Pauli basis, and* $\lvert\Psi\rangle$ *is an eigenstate of the third Pauli operator. This conclusion can be checked directly.*

**Example 5** (Nonclassical KD distribution that violates $2N_A + 2N_F > 3d + n_\parallel - 3\bar{n}_\parallel$)**.** *Satisfying* $2N_A + 2N_F > 3d + n_\parallel - 3\bar{n}_\parallel$ *suffices to guarantee a nonclassical KD distribution. But it is not necessary, as we demonstrate here. Consider a two-qubit system. We choose* $\hat{A}$ *and* $\hat{F}$ *such that* $\{\lvert a_i\rangle\} = \{\lvert 0\rangle\lvert 0\rangle, \lvert 0\rangle\lvert 1\rangle, \lvert 1\rangle\lvert 0\rangle, \lvert 1\rangle\lvert 1\rangle\}$ *and* $\{\lvert f_j\rangle\} = \{\lvert 0\rangle\lvert +\rangle, \lvert 0\rangle\lvert -\rangle, \lvert 1\rangle\lvert +\rangle, \lvert 1\rangle\lvert -\rangle\}$*. We set* $\hat{\rho} = \lvert\Psi\rangle\langle\Psi\rvert$*, where* $\lvert\Psi\rangle = (\lvert 0\rangle\lvert 0\rangle + 2\lvert 0\rangle\lvert 1\rangle)/\sqrt{5}$*. These choices imply* $N_A = N_F = 2$*,* $d = 4$*, and* $n_\parallel = \bar{n}_\parallel = 0$*. Hence the inequality above is violated:* $8 \not> 12$*. Nonetheless, the KD distribution is nonclassical (Table V).*

TABLE V. The KD distribution of Ex. 5.

| $\lvert f_j\rangle$ \ $\lvert a_i\rangle$ | $\lvert 0\rangle\lvert 0\rangle$ | $\lvert 0\rangle\lvert 1\rangle$ | $\lvert 1\rangle\lvert 0\rangle$ | $\lvert 1\rangle\lvert 1\rangle$ |
|---|---|---|---|---|
| $\lvert 0\rangle\lvert +\rangle$ | $\frac{3}{10}$ | $\frac{3}{5}$ | 0 | 0 |
| $\lvert 0\rangle\lvert -\rangle$ | $-\frac{1}{10}$ | $\frac{1}{5}$ | 0 | 0 |
| $\lvert 1\rangle\lvert +\rangle$ | 0 | 0 | 0 | 0 |
| $\lvert 1\rangle\lvert -\rangle$ | 0 | 0 | 0 | 0 |

## Appendix B: Proof of Thm. 1

For convenience, we first assume that no $|a_i\rangle$ and $|f_j\rangle$ are parallel: $n_\parallel = \bar{n}_\parallel = 0$. Then, we generalize.

Assume that the KD distribution is classical: $q_{i,j}^{\hat{\rho}} \in \mathbb{R}_{\geq 0}$ for all $i, j$. Without changing the quasiprobabilities or the observables, we can redefine the vectors through $|a_i\rangle \mapsto e^{i\alpha_i} |a_i\rangle$ and $|f_j\rangle \mapsto e^{i\phi_j} |f_j\rangle$. We choose the $\alpha_i, \phi_j \in \mathbb{R}$ such that $\langle a_i|\Psi\rangle \langle\Psi|f_j\rangle \in \mathbb{R}_{\geq 0}$. By assumption, $\langle f_j|a_i\rangle \langle a_i|\Psi\rangle \langle\Psi|f_j\rangle \in \mathbb{R}_{\geq 0}$. Hence, for each $i$ and $j$, $\langle a_i|f_j\rangle \in \mathbb{R}_{\geq 0}$, or $\langle a_i|\Psi\rangle = 0$, or $\langle\Psi|f_j\rangle = 0$. Let $\hat{U}$ denote the unitary operator that rotates $\mathcal{V}_A$ into $\mathcal{V}_F$. $\hat{U}$ is represented, relative to $\mathcal{V}_F$, by the matrix with elements $\hat{U}_{i,j} = \langle a_i|f_j\rangle$. $d - N_A$ vectors in $\mathcal{V}_A$, and $d - N_F$ vectors in $\mathcal{V}_F$, are orthogonal to $|\Psi\rangle$. Hence, at most $d - N_A$ rows and $d - N_F$ columns of $\hat{U}$ contain negative or nonreal values.

Let us order $\mathcal{V}_A$ and $\mathcal{V}_F$ so that the top left-hand $N_A$-by-$N_F$ block contains only non-negative real entries (Fig. 1). The top $N_A$ entries of each column $j$ form a "top vector" $\mathbf{f}_j^{\rm t} \in \mathbb{R}_{\geq 0}^{N_A}$. The bottom $d - N_A$ entries of column $j$ form a "bottom vector" $\mathbf{f}_j^{\rm b} \in \mathbb{C}^{d-N_A}$. We label columns 1 to $k$ "left," columns $k+1$ to $N_F$ "middle," and columns $N_F + 1$ to $d$ "right."

For $j = 1, 2, \ldots, N_F$, all elements of each $\mathbf{f}_j^{\rm t}$ are non-negative reals. Hence $(\mathbf{f}_\ell^{\rm t})^\top \mathbf{f}_m^{\rm t} \geq 0$ for all $\ell, m \in \{1, \ldots, N_F\}$. Therefore, for the columns of $\hat{U}$ to be orthogonal, $(\mathbf{f}_\ell^{\rm b})^\dagger \mathbf{f}_m^{\rm b} \leq 0$ must hold for all $\ell, m \in \{1, \ldots, N_F\}$ for which $\ell \neq m$. This inner-product constraint implies the following lemma.

**Lemma 1.** *At most $2(d - N_A)$ of the $N_F$ left and middle bottom vectors are nonzero.*

*Proof of Lem. 1:* Here, we bound the maximum number of nonzero bottom vectors whose pairwise products are $\leq 0$. Let $\mathcal{S} = \{\mathbf{s}_j\}$ denote a set of nonzero vectors in $\mathbb{C}^n$ whose pairwise inner products are $\leq 0$. We use an orthonormal basis in terms of which $\mathbf{s}_1 \to (s_1, 0, \ldots, 0)^\top$ and $s_1 > 0$. Every other vector $\mathbf{s}_j \in \mathcal{S} \setminus \{\mathbf{s}_1\}$ is represented by a column with first element $\leq 0$. Hence, for these other $\mathbf{s}_j$ to have inner products $\leq 0$, the vectors formed from their last $n-1$ entries must all have inner products $\leq 0$. At most one of these shorter vectors can be the null vector $\mathbf{0}$. So all the others are nonzero vectors in $\mathbb{C}^{n-1}$ whose pairwise inner products are $\leq 0$. The relevant vectors space's dimensionality has decreased to $n-1$. Proceeding from $n$ to $n-1$, we have "lost" at most two vectors, $\mathbf{s}_1 \to (s_1, 0, \ldots, 0)^\top$ and $\mathbf{s}_2 \to (-s_2, 0, \ldots, 0)^\top$, where $s_1, s_2 > 0$. By induction, $\mathcal{S}$ can have at most $2n$ vectors. In the proof of Thm. 1, $n = d - N_A$. Consequently, $\leq 2(d - N_A)$ of the $N_F$ left and middle bottom vectors are nonzero.□

Lemma 1 ensures that if $k$ denotes the number of nonzero elements of $\{\mathbf{f}_1^{\rm b}, \ldots, \mathbf{f}_{N_F}^{\rm b}\}$, then

$$k \leq 2(d - N_A). \tag{B1}$$

Let us order the columns of $\hat{U}$ so that the $k$ nonzero bottom vectors occupy columns 1 to $k$, while $\mathbf{f}_{k+1}^{\rm b} = \mathbf{f}_{k+2}^{\rm b} = \ldots = \mathbf{f}_{N_F}^{\rm b} = \mathbf{0}$. (Fig. 1).

Columns 1 to $k$ (the left columns) are linearly independent. Therefore, the collection of columns contains nonzero entries in $\geq k$ rows. Up to $d - N_A$ of those rows can be in the bottom vectors (which contain exactly $d - N_A$ rows). The left top vectors make up the difference, having nonzero entries in $\geq k - (d - N_A)$ rows. The middle top vectors must contain only 0s in these rows, since they are orthogonal to the left top vectors.[7] Let us order the rows of $\hat{U}$ such that the middle top vectors' uppermost $\geq k - (d - N_A)$ entries are 0s (Fig. 1). Only the middle top vectors' lower $\leq N_A - [k - (d - N_A)] = d - k$ entries can be nonzero. By assumption, no $|a_i\rangle$ is parallel to any $|f_j\rangle$. So each middle top vector has $\geq 2$ nonzero entries $\langle a_i|f_j\rangle$. But the middle top vectors are mutually orthogonal, and all their entries

[7] The middle columns are orthogonal to the left columns. The middle bottom columns' being **0**s forces the middle top vectors to be orthogonal to the left top vectors.

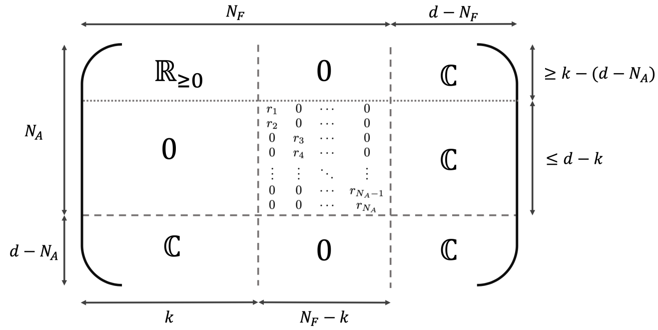


FIG. 1. Unitary matrix with entries $\hat{U}_{i,j} = \langle a_i | f_j \rangle$. The dashed vertical lines divide the columns into "left," "middle," and "right" sets. The dashed horizontal line divides the rows into "top" and "bottom" sets. The vectors $|a_i\rangle$ and $|f_j\rangle$ are ordered such that any nonreal or negative $\hat{U}_{i,j}$ appear in the bottom rows or rightmost columns.

$\geq 0$. So no two middle top vectors can have nonzero elements in the same row. Therefore, $2(N_F - k) \leq d - k$. We bound $k$ with Ineq. B1 and rearrange:

$$2N_A + 2N_F \leq 3d. \tag{B2}$$

Finally, we extend $2N_F + 2N_A \leq 3d$ [Ineq. (B2)] to scenarios in which $\bar{n}_\parallel \neq 0$ or $n_\parallel \neq 0$, completing the proof of Thm. 1. We first remove any pairs $(|a_i\rangle, |f_j\rangle)$ of parallel vectors from $\mathcal{V}_A$ and $\mathcal{V}_F$. Consider the subspace $\mathcal{H}'$ spanned by the remaining basis vectors. Let $d' \equiv \dim(\mathcal{H}')$. Define $N'_A$ as the number of $|a_i\rangle$ that have nonzero overlaps with $|\Psi\rangle$, and define $N'_F$ analogously. Denote by $|\Psi'\rangle$ the projection of $|\Psi\rangle$ onto $\mathcal{H}'$. Inequality (B2) can be rederived for this reduced subspace: $2N'_F + 2N'_A \leq 3d'$. (If $|\Psi'\rangle = \mathbf{0}$, then $N'_A = N'_F = 0$, so the inequality still holds.) Substituting in from $N'_A + n_\parallel = N_A$, $N'_F + n_\parallel = N_F$, and $d' + n_\parallel + \bar{n}_\parallel = d$ leads to $2N_F + 2N_A \leq 3d + n_\parallel - 3\bar{n}_\parallel$.

We derived this inequality assuming a classical KD distribution. A violation of the inequality implies nonclassicality.□

### Appendix C: Properties of the imaginary and real components of the KD distribution

Consider an experiment that involves eigenbases $\{|a_i\rangle\}$ and $\{|f_j\rangle\}$ or, equivalently, nondegenerate operators $\hat{A}$ and $\hat{F}$. One might want to construct a KD distribution $\{q^{\hat{\rho}}_{i,j}\}$ that has, or that lacks, KD nonclassicality by picking a suitable $\hat{\rho}$. Furthermore, one might want specific quasiprobabilities $q^{\hat{\rho}}_{i,j}$ to have negative or nonreal nonclassicality. We provide useful results for tailoring $\hat{\rho}$.

As in part of the main text, we assume that $\hat{\rho}$ is pure: $\hat{\rho} = |\Psi\rangle\langle\Psi|$. The imaginary part of $q^{\hat{\rho}}_{i,j}$ decomposes as

$$\Im\left[q^{\hat{\rho}}_{i,j}\right] = \frac{1}{2i}\left[q^{\hat{\rho}}_{i,j} - \left(q^{\hat{\rho}}_{i,j}\right)^*\right] = \frac{1}{2}\mathrm{Tr}\left[\hat{H}_{i,j}\hat{\rho}\right], \tag{C1}$$

where $\hat{H}_{i,j} \equiv i\hat{\Pi}^a_i\hat{\Pi}^f_j - i\hat{\Pi}^f_j\hat{\Pi}^a_i$. If $|\langle a_i|f_j\rangle| \neq 0, 1$, then $-i\hat{H}_{i,j}$ is the antisymmetric product of two noncommuting rank-1 projectors. Under this condition, $\hat{H}_{i,j}$ also has two eigenvalues, $h^{(\pm)}_{i,j} = \pm|\langle a_i|f_j\rangle|\sqrt{1 - |\langle a_i|f_j\rangle|^2} \neq 0$, with respective eigenvectors

$$|h^{(\pm)}_{i,j}\rangle = \frac{1}{\sqrt{2}}\left[\left(\mp 1 - i\frac{|\langle a_i|f_j\rangle|}{\sqrt{1 - |\langle a_i|f_j\rangle|^2}}\right)e^{i\mathrm{Arg}(\langle a_i|f_j\rangle)}|a_i\rangle + i\frac{1}{\sqrt{1 - |\langle a_i|f_j\rangle|^2}}|f_j\rangle\right]. \tag{C2}$$

The real part of $q^{\hat{\rho}}_{i,j}$ can be written as

$$\Re\left[q^{\hat{\rho}}_{i,j}\right] = \frac{1}{2}\mathrm{Tr}\left[q^{\hat{\rho}}_{i,j} + \left(q^{\hat{\rho}}_{i,j}\right)^*\right] \equiv \frac{1}{2}\mathrm{Tr}\left[\hat{G}_{i,j}\hat{\rho}\right], \tag{C3}$$

where $\hat{G}_{i,j} \equiv \hat{\Pi}^a_i\hat{\Pi}^f_j + \hat{\Pi}^f_j\hat{\Pi}^a_i$. If $|\langle a_i|f_j\rangle| \neq 0$, then $\hat{G}_{i,j}$ is the symmetric product of two noncommuting rank-1 projectors. Under this condition, $\hat{G}_{i,j}$ also has two eigenvalues, $g^{(\pm)}_{i,j} = |\langle a_i|f_j\rangle|(|\langle a_i|f_j\rangle| \pm 1) \neq 0$, with corresponding eigenvectors

$$|g^{(\pm)}_{i,j}\rangle = \frac{1}{\sqrt{2}}\left(|f_j\rangle \pm e^{i\mathrm{Arg}(\langle a_i|f_j\rangle)}|a_i\rangle\right). \tag{C4}$$

$g_{i,j}^{(+)}$ and $g_{i,j}^{(-)}$ are positive and negative, respectively.

Given the eigenvalues $h_{i,j}^{(\pm)}$ and $g_{i,j}^{(\pm)}$, and the eigenvectors $|h_{i,j}^{(\pm)}\rangle$ and $|g_{i,j}^{(\pm)}\rangle$, one can tailor $|\Psi\rangle$ such that a quasiprobability $q_{i,j}^{\hat{\rho}}$ has a negative real component, or an imaginary component, of a certain magnitude.

### Appendix D: Extension to restricted information, or coarse-grained KD distributions

$\hat{A}$ can be degenerate, as can $\hat{F}$. Regardless, $\hat{A}$ eigendecomposes as $\hat{A} = \sum_l A_l \hat{A}_l$, where $\hat{A}_l \equiv \sum_{i\,:\,|a_i\rangle \in \mathcal{A}_l} |a_i\rangle \langle a_i|$ and $\mathcal{A}_l$ is the eigensubspace associated with the eigenvalue $A_l$. Similarly, $\hat{F} = \sum_k F_k \hat{F}_k$, where $\hat{F}_k \equiv \sum_{j\,:\,|f_j\rangle \in \mathcal{F}_k} |f_j\rangle \langle f_j|$ and $\mathcal{F}_k$ is eigensubspace associated with the eigenvalue $F_k$. If any $\hat{F}_k$ ($\hat{A}_l$) has rank $> 1$, $\hat{F}_k$ ($\hat{A}_l$) has nonequivalent eigenbases. Consequently, $\{q_{i,j}^{\hat{\rho}}\}$ is generally not unique for a fixed $\hat{\rho}$. This degeneracy problem arises in, e.g., studies of quantum scrambling: $\hat{A}$ and $\hat{F}$ manifest as local observables of a many-body system and so are degenerate [38, 51–53]. We therefore define a *coarse-grained KD quasiprobability distribution* by marginalizing $\{q_{i,j}^{\hat{\rho}}\}$ over the degeneracies:

$$\left\{ \mathcal{Q}_{l,k}^{\hat{\rho}} \right\} := \left\{ \sum_{\substack{i\,:\,|a_i\rangle \in \mathcal{A}_l \\ j\,:\,|f_j\rangle \in \mathcal{F}_k}} \langle f_j | a_i \rangle \langle a_i | \hat{\rho} | f_j \rangle \right\} = \left\{ \mathrm{Tr}\left( \hat{F}_k \hat{A}_l \hat{\rho} \right) \right\}. \tag{D1}$$

The projectors $\hat{F}_k$ and $\hat{A}_l$ are unique. So, for a given $\hat{\rho}$, the quasiprobabilities $\mathcal{Q}_{l,k}^{\hat{\rho}}$ are unique.

We now prove a theorem analogous to Thm. 1 for the coarse-grained distribution, providing a necessary condition for $\{\mathcal{Q}_{l,k}^{\hat{\rho}}\}$ to be classical when $\hat{\rho} = |\Psi\rangle \langle \Psi|$ is pure. In analogy with Eq. (4), we define as $\tilde{N}_A$ the number of $\hat{A}$ eigenspaces onto which $|\Psi\rangle$ has nonzero projections. In analogy with Eq. (5), we define $\tilde{N}_F$ similarly:

$$\tilde{N}_A \equiv ||\{l : \hat{A}_l |\Psi\rangle \neq 0\}||, \text{ and} \tag{D2}$$

$$\tilde{N}_F \equiv ||\{k : \hat{F}_k |\Psi\rangle \neq 0\}||. \tag{D3}$$

In analogy with previous definitions, we denote by $\tilde{n}_{\parallel}$ (respectively, $\tilde{\bar{n}}_{\parallel}$) the number of $\hat{A}_l |\Psi\rangle$ that are (i) parallel to some $\hat{F}_k |\Psi\rangle$ and (ii) nonorthogonal (respectively, orthogonal) to $|\Psi\rangle$. This background informs the following theorem, which resembles Thm. 1.

**Theorem 3** (Sufficient conditions for coarse-grained Kirkwood-Dirac nonclassicality)**.** *Suppose that $\hat{\rho} = |\Psi\rangle \langle \Psi|$ is pure. If $2\tilde{N}_F + 2\tilde{N}_A > 3d + \tilde{n}_{\parallel} - 3\tilde{\bar{n}}_{\parallel}$, the coarse-grained KD distribution is nonclassical.*

*Proof:* As in the proof of Thm. 1, we begin by assuming that the KD distribution is classical: $\mathcal{Q}_{l,k}^{\hat{\rho}} \geq 0$ for all $l, k$. We assume that $\tilde{n}_{\parallel} = \tilde{\bar{n}}_{\parallel} = 0$; then, we generalize.

Define the $\tilde{N}_A$ nonzero projections $|a_l^{\Psi}\rangle \equiv \hat{A}_l |\Psi\rangle / ||\hat{A}_l |\Psi\rangle||$ and the $\tilde{N}_F$ nonzero projections $|f_k^{\Psi}\rangle \equiv \hat{F}_k |\Psi\rangle / ||\hat{F}_k |\Psi\rangle||$. By appending vectors to the sets $\{|a_l^{\Psi}\rangle\}$ and $\{|f_k^{\Psi}\rangle\}$, we can form orthonormal bases $\mathcal{B}_A$ and $\mathcal{B}_F$. By the sets' definitions, $|\Psi\rangle \in \mathrm{span}\{|a_l^{\Psi}\rangle\}$, and $|\Psi\rangle \in \mathrm{span}\{|f_k^{\Psi}\rangle\}$. Therefore, the appended vectors are orthogonal to $|\Psi\rangle$. Since $\mathcal{Q}_{l,k}^{\hat{\rho}} = \mathrm{Tr}\left(\hat{F}_k \hat{A}_l \hat{\rho}\right) = \langle \Psi | \hat{F}_k \hat{A}_l |\Psi\rangle = \langle f_k^{\Psi} | a_l^{\Psi} \rangle \times ||\hat{A}_l |\Psi\rangle|| \times ||\hat{F}_k |\Psi\rangle||$, the condition $\mathcal{Q}_{l,k}^{\hat{\rho}} \geq 0$ implies that $\langle f_k^{\Psi} | a_l^{\Psi} \rangle \geq 0$. Therefore, any nonclassical quasiprobabilities contain vectors appended to the bases $\mathcal{B}_A$ and $\mathcal{B}_F$. But the appended basis elements are orthogonal to $|\Psi\rangle$ and so appear only in zero-valued quasiprobabilities. Therefore, $\mathcal{B}_A$ and $\mathcal{B}_F$ define a classical non-coarse-grained KD distribution for $\hat{\rho}$. Let this non-coarse-grained KD distribution's $N_A, N_F, n_{\parallel}$ and $\bar{n}_{\parallel}$ be defined as in the proof of Thm. 1. By Thm. 1, $2N_A + 2N_F \leq 3d + n_{\parallel} - 3\bar{n}_{\parallel}$. Since we extended the bases with vectors orthogonal to $|\Psi\rangle$, $N_A = \tilde{N}_A$, $N_F = \tilde{N}_F$, and $n_{\parallel} = \tilde{n}_{\parallel} = 0$. Therefore, $2\tilde{N}_A + 2\tilde{N}_F \leq 3d - \bar{n}_{\parallel} \leq 3d$. The generalization to $\tilde{n}_{\parallel} \neq 0$ or $\tilde{\bar{n}}_{\parallel} \neq 0$ proceeds as in App. **??**. Therefore, every classical coarse-grained KD distribution satisfies $2\tilde{N}_F + 2\tilde{N}_A \leq 3d + \tilde{n}_{\parallel} - 3\tilde{\bar{n}}_{\parallel}$. Violating this inequality suffices for the coarse-grained distribution to be nonclassical. □

An analog of Cor. 1 follows.

**Corollary 2.** *Suppose that at least one of $\hat{A}$ and $\hat{F}$ is nondegenerate, while the other is not completely degenerate. If the KD distribution lacks zero-valued quasiprobabilities, $\mathcal{Q}_{l,k}^{\hat{\rho}}$ is nonclassical.*

*Proof:* Suppose that all the $\mathcal{Q}^{\hat{\rho}}_{l,k}$ are nonzero. Without loss of generality, assume that $\hat{A}$ is nondegenerate. $\hat{F}$ is not completely degenerate, so its eigendecomposition contains at least two distinct projectors, $\hat{F}_1$ and $\hat{F}_2$. Since the $\mathcal{Q}^{\hat{\rho}}_{l,k}$ are nonzero, $\hat{F}_1 |\Psi\rangle$ and $\hat{F}_2 |\Psi\rangle$ are nonzero, by Eq. (D1). Therefore, there exist at least two vectors, $|f_1^\Psi\rangle$ and $|f_2^\Psi\rangle$, as defined in the proof of Thm. 3.

The rest of the proof is a proof by contradiction. Suppose that $\{\mathcal{Q}^{\hat{\rho}}_{l,k}\}$ is classical. If it lacks zero-valued quasiprobabilities, then $\langle f_k^\Psi | a_l\rangle \in \mathbb{R}_{>0}$ for every $l$ and $k$. By the $\hat{F}$ eigenspaces' orthogonality,

$$0 = \langle f_1^\Psi | f_2^\Psi \rangle = \sum_l \langle f_1^\Psi | a_l\rangle \langle a_l | f_2^\Psi\rangle > 0. \tag{D4}$$

The final inequality follows because $\langle f_1^\Psi | a_l\rangle, \langle a_l | f_2^\Psi\rangle > 0$ for each $l$. Implying the contradiction $0 > 0$, the assumption of the distribution's classicality is false. □

Let us briefly discuss the case, consistent with the assumptions of Cor. 2, in which $\hat{F}$ is degenerate and $\hat{A}$ is not (or *vice versa*). Coarse-graining over one index suffices to define a unique KD distribution distribution:

$$\left\{Q^{\hat{\rho}}_{i,k}\right\} := \left\{ \sum_{j\,:\,|f_j\rangle \in \mathcal{F}_k} \langle f_j | a_i\rangle \langle a_i | \hat{\rho} | f_j\rangle \right\} = \left\{ \mathrm{Tr}\left( \hat{F}_k |a_i\rangle \langle a_i | \hat{\rho} \right) \right\}. \tag{D5}$$

Such a distribution has been used, for example, in postselected quantum metrology. In Ref. [56], $\hat{F} = 0 \times \sum_{j\,:\,|f_j\rangle\in\mathcal{F}_0} |f_j\rangle\langle f_j| + 1 \times \sum_{j'\,:\,|f'_j\rangle\in\mathcal{F}_1} |f'_j\rangle\langle f'_j|$ is an observable whose measured value determines whether a quantum state should be discarded or funnelled to further processing. If the coarse-grained KD distribution contains negative values, a metrological protocol may provide a nonclassical advantage. Further properties of $Q^{\hat{\rho}}_{i,k}$ are proved below.

### 1. Properties of the imaginary and real components of the coarse-grained KD distribution

Here, we extend the results of App. C to $\{Q^{\hat{\rho}}_{i,k}\}$. Suppose that $\hat{\rho}$ is pure: $\hat{\rho} = |\Psi\rangle\langle\Psi|$. The imaginary part of $Q^{\hat{\rho}}_{i,k}$ decomposes as

$$\Im\left[Q^{\hat{\rho}}_{i,k}\right] = \frac{1}{2i}\left[Q^{\hat{\rho}}_{i,k} - \left(Q^{\hat{\rho}}_{i,k}\right)^*\right] = \frac{1}{2}\mathrm{Tr}\left[\hat{R}_{i,k}\hat{\rho}\right], \tag{D6}$$

where $\hat{R}_{i,k} \equiv i\hat{\Pi}^a_i \hat{F}_k - i\hat{F}_k\hat{\Pi}^a_i$. If $p^a_F \equiv \mathrm{Tr}\left(\hat{\Pi}^a_i\hat{F}_k\right) \neq 0, 1$, then $\hat{R}_{i,k}$ has two nonzero eigenvalues, $r^{(\pm)}_{i,k} = \pm\sqrt{p^a_F - (p^a_F)^2}$. The eigenvectors are

$$|r^{(\pm)}_{i,k}\rangle = \frac{1}{\sqrt{2}}\left[\left(\mp\frac{1}{\sqrt{p^a_F}} + i\frac{1}{\sqrt{1-p^a_F}}\right)\hat{F}_k|a_i\rangle - i\frac{1}{\sqrt{1-p^a_F}}|a_i\rangle\right]. \tag{D7}$$

Similarly, the real part of $Q^{\hat{\rho}}_{i,k}$ can be expressed as

$$\Re\left[Q^{\hat{\rho}}_{i,k}\right] = \frac{1}{2}\left[Q^{\hat{\rho}}_{i,k} + \left(Q^{\hat{\rho}}_{i,k}\right)^*\right] = \frac{1}{2}\mathrm{Tr}\left[\hat{S}_{i,k}\hat{\rho}\right], \tag{D8}$$

where $\hat{S}_{i,k} \equiv \hat{\Pi}^a_i\hat{F}_k + \hat{F}_k\hat{\Pi}^a_i$. If $p^a_F \neq 0, 1$, then $\hat{S}_{i,k}$ has two eigenvalues, $s^{(\pm)}_{i,k} = p^a_F \pm \sqrt{p^a_F}$. The eigenvectors are

$$|s^{(\pm)}_{i,k}\rangle = \frac{1}{\sqrt{2}}\left[|a_i\rangle \pm \frac{1}{\sqrt{p^a_F}}\hat{F}_k|a_i\rangle\right]. \tag{D9}$$

## Appendix E: Proof of Thm. 2

Here, we upper-bound $\mathcal{N}\left(\{q^{\hat{\rho}}_{i_1,\ldots,i_k}\}\right)$, proving Thm. 2. First, we restrict our attention pure states $\hat{\rho} = |\Psi\rangle\langle\Psi|$. We prove that $\mathcal{N}\left(\{q^{|\Psi\rangle\langle\Psi|}_{i_1,\ldots,i_k}\}\right)$ maximizes when each of its inner products has magnitude $1/\sqrt{d}$. Thus, if $\mathcal{N}\left(\{q^{|\Psi\rangle\langle\Psi|}_{i_1,\ldots,i_k}\}\right)$

is maximized, then $|\langle a^{(1)}_{i_1}|\Psi\rangle| = |\langle a^{(k)}_{i_k}|\Psi\rangle| = \frac{1}{\sqrt{d}}$ for all $i_1, i_k$. Every $\hat{\rho}$ equals a convex sum of pure states $\hat{\rho}_n$. By the triangle inequality, $\mathcal{N}\left(\{q^{\hat{\rho}}_{i_1,\ldots,i_k}\}\right)$ is upper-bounded by a convex sum of the $\mathcal{N}\left(\{q^{\hat{\rho}_n}_{i_1,\ldots,i_k}\}\right)$. Therefore, at any maximum of $\mathcal{N}\left(\{q^{\hat{\rho}}_{i_1,\ldots,i_k}\}\right)$, $\hat{\rho}$ is a linear combination of pure states, each of which maximizes $\mathcal{N}\left(\{q^{\hat{\rho}}_{i_1,\ldots,i_k}\}\right)$. We finish the proof by showing that no such mixed state maximizes $\mathcal{N}\left(\{q^{\hat{\rho}}_{i_1,\ldots,i_k}\}\right)$. Hence, only pure states that are unbiased with respect to $\hat{A}_1$ and $\hat{A}_k$ eigenbases, as described above, maximize $\mathcal{N}\left(\{q^{\hat{\rho}}_{i_1,\ldots,i_k}\}\right)$.

Our proof requires the following lemma:

**Lemma 1.** *Let $\{|i\rangle\}_{i=1}^d$ be an orthonormal basis for a $d$-dimensional Hilbert space $\mathcal{H}$. The unit vector $|\psi\rangle \in \mathcal{H}$ satisfies $\sum_{i=1}^d |\langle i|\psi\rangle| \leq \sqrt{d}$. The bound is saturated if and only if $|\langle i|\psi\rangle| = \frac{1}{\sqrt{d}}$ for every $i$.*

*Proof:* By Jensen's inequality,

$$\left(\sum_{i=1}^d |\langle i|\psi\rangle|\right)^2 \leq d\sum_{i=1}^d |\langle i|\psi\rangle|^2 = d. \tag{E1}$$

Comparing the first and third expressions, we conclude that

$$\sum_{i=1}^d |\langle i|\psi\rangle| \leq \sqrt{d}. \tag{E2}$$

Jensen's inequality is saturated if and only if the terms in the first sum in (E1) equal each other, as can be inferred from the geometric proof of Jensen's inequality. Consequently, Ineq. (E2) is saturated if and only if $|\langle i|\psi\rangle| = 1/\sqrt{d}$. □

To upper-bound $\mathcal{N}\left(\{q^{\hat{\rho}}_{i_1,\ldots,i_k}\}\right)$, we assume that $\rho = |\Psi\rangle\langle\Psi|$ is pure. By Eqs. (2) and (6),

$$\mathcal{N}\left(\left\{q^{|\Psi\rangle\langle\Psi|}_{i_1,\ldots,i_k}\right\}\right) = -1 + \sum_{i_1,\ldots,i_k} |\langle a^{(1)}_{i_1}|a^{(2)}_{i_2}\rangle \times \ldots \times \langle a^{(k)}_{i_k}|\Psi\rangle\langle\Psi|a^{(1)}_{i_1}\rangle|, \tag{E3}$$

where $\{|a^{(n)}_{i_n}\rangle\}_{i_n=1}^d$ is an eigenbasis of Hermitian operator $A^{(n)}$. [To simplify notation in this proof, we have labeled operators differently than in Eq. (2): Here, the $\langle a^{(k)}_{i_k}|$ acts on $|\Psi\rangle$.] We now show that the RHS of Eq. (E3) maximizes when the magnitude of all the inner products in $\mathcal{N}\left(\{q^{|\Psi\rangle\langle\Psi|}_{i_1,\ldots,i_k}\}\right)$ equal each other.

For a fixed value of $i_1$,

$$\sum_{i_2,\ldots,i_k} |\langle a^{(1)}_{i_1}|a^{(2)}_{i_2}\rangle \times \ldots \times \langle a^{(k)}_{i_k}|\Psi\rangle\langle\Psi|a^{(1)}_{i_1}\rangle| = \sum_{i_2}\left(|\langle a^{(1)}_{i_1}|a^{(2)}_{i_2}\rangle| \times \sum_{i_3,\ldots,i_k} |\langle a^{(2)}_{i_2}|a^{(3)}_{i_3}\rangle \times \ldots \times \langle a^{(k)}_{i_k}|\Psi\rangle\langle\Psi|a^{(1)}_{i_1}\rangle|\right) \tag{E4}$$

$$\leq \sum_{i_2} |\langle a^{(1)}_{i_1}|a^{(2)}_{i_2}\rangle| \times \max_{i'_2} \sum_{i_3,\ldots,i_k} |\langle a^{(2)}_{i'_2}|a^{(3)}_{i_3}\rangle \times \ldots \times \langle a^{(k)}_{i_k}|\Psi\rangle\langle\Psi|a^{(1)}_{i_1}\rangle| \tag{E5}$$

$$\leq \sqrt{d} \times \max_{i'_2} \sum_{i_3,\ldots,i_k} |\langle a^{(2)}_{i'_2}|a^{(3)}_{i_3}\rangle \times \ldots \times \langle a^{(k)}_{i_k}|\Psi\rangle\langle\Psi|a^{(1)}_{i_1}\rangle|. \tag{E6}$$

Inequality (E5) follows because, if $x_j$ and $y_j$ are non-negative real numbers, then $\sum_j x_j y_j \leq \sum_j x_j \times \max_{j'} y_{j'}$. Inequality (E6) follows from Lemma 1. Proceeding from the left-hand side of Eq. (E4) to the RHS of (E6), we (i) reduce the number of summed indices by 1 and (ii) acquire a factor of $\sqrt{d}$. Let us iterate this step $k-3$ more times:

$$\sum_{i_2,\ldots,i_k} |\langle a^{(1)}_{i_1}|a^{(2)}_{i_2}\rangle \times \ldots \times \langle a^{(k)}_{i_k}|\Psi\rangle\langle\Psi|a^{(1)}_{i_1}\rangle| \leq (\sqrt{d})^2 \times \max_{i'_3} \sum_{i_4,\ldots,i_k} |\langle a^{(3)}_{i'_3}|a^{(4)}_{i_4}\rangle \times \ldots \times \langle a^{(k)}_{i_k}|\Psi\rangle\langle\Psi|a^{(1)}_{i_1}\rangle| \tag{E7}$$

$$\leq \ldots \tag{E8}$$

$$\leq (\sqrt{d})^{k-2} \times \max_{i'_{k-1}} \sum_{i_k} |\langle a^{(k-1)}_{i'_{k-1}}|a^{(k)}_{i_k}\rangle\langle a^{(k)}_{i_k}|\Psi\rangle\langle\Psi|a^{(1)}_{i_1}\rangle|. \tag{E9}$$

Summing over $i_1$ yields

$$\sum_{i_1,\ldots,i_k} |\langle a^{(1)}_{i_1}|a^{(2)}_{i_2}\rangle \times \ldots \times \langle a^{(k)}_{i_k}|\Psi\rangle \langle \Psi|a^{(1)}_{i_1}\rangle| \leq (\sqrt{d})^{k-2} \times \max_{i'_{k-1}} \sum_{i_1,i_k} |\langle a^{(k-1)}_{i'_{k-1}}|a_{i_k}\rangle \langle a^{(k)}_{i_k}|\Psi\rangle \langle \Psi|a^{(1)}_{i_1}\rangle| \tag{E10}$$

$$= d^{\frac{k}{2}-1} \sum_{i_1} |\langle \Psi|a^{(1)}_{i_1}\rangle| \times \max_{i'_{k-1}} \sum_{i_k} |\langle a^{(k-1)}_{i'_{k-1}}|a^{(k)}_{i_k}\rangle \langle a^{(k)}_{i_k}|\Psi\rangle| \tag{E11}$$

$$\leq d^{\frac{k-1}{2}} \times \max_{i'_{k-1}} \sum_{i_k} |\langle a^{(k-1)}_{i'_{k-1}}|a^{(k)}_{i_k}\rangle| \times |\langle a^{(k)}_{i_k}|\Psi\rangle| \tag{E12}$$

$$\leq d^{\frac{k-1}{2}} \times \max_{i'_{k-1}} \sqrt{\sum_{i_k} |\langle a^{(k-1)}_{i'_{k-1}}|a^{(k)}_{i_k}\rangle|^2 \times \sum_{i'_k} |\langle a^{(k)}_{i'_k}|\Psi\rangle|^2} \tag{E13}$$

$$= d^{\frac{1}{2}(k-1)}. \tag{E14}$$

Inequality (E12) follows from Lemma 1. Inequality (E13) follows from the Cauchy-Schwarz inequality: For vectors $\vec{u}, \vec{v} \in \mathbb{R}^n$, denote the inner product by $(\vec{u}, \vec{v}) = \sum_{j=1}^d u_j v_j$. The Cauchy-Schwarz inequality implies that $(\vec{u}, \vec{v})^2 \leq (\vec{u}, \vec{u})(\vec{v}, \vec{v})$. Let $\vec{u} = \left(|\langle a^{(k-1)}_{i'_{k-1}}|a^{(k)}_1\rangle|, |\langle a^{(k-1)}_{i'_{k-1}}|a^{(k)}_2\rangle|, \ldots, |\langle a^{(k-1)}_{i'_{k-1}}|a^{(k)}_d\rangle|\right)$ and $\vec{v} = \left(|\langle a^{(k)}_1|\Psi\rangle|, |\langle a^{(k)}_2|\Psi\rangle|, \ldots, |\langle a^{(k)}_d|\Psi\rangle|\right)$. Square-rooting each side of the Cauchy-Schwarz inequality yields Ineq. (E13). Therefore,

$$\mathcal{N}\left(\{q^{\hat{\rho}}_{i_1,\ldots,i_k}\}\right) \leq d^{(k-1)/2} - 1. \tag{E15}$$

It is easy to see that, if all the inner products in $\{q^{\hat{\rho}}_{i_1,\ldots,i_k}\}$ have magnitudes $1/\sqrt{d}$, Ineq. (E15) is saturated. This criterion is satisfied when two conditions hold simultaneously: (i) $\hat{A}^{(i)}$ and $\hat{A}^{(i+1)}$ have mutually unbiased eigenbases for each $i = 1, 2, \ldots, k-1$; and (ii) $|\langle a^{(1)}_{i_1}|\Psi\rangle| = |\langle a^{(k)}_{i_k}|\Psi\rangle| = \frac{1}{\sqrt{d}}$ for all $i_1, i_k$.

These two conditions are not only sufficient, but also necessary for $\mathcal{N}\left(\{q^{\hat{\rho}}_{i_1,\ldots,i_k}\}\right)$ to be maximized: Inequalities (E5)-(E9) are all saturated only if (i) holds. Inequalities (E12) and (E13) are saturated only if (ii) holds.

Therefore, if a (possibly mixed) state $\hat{\rho}$ maximizes $\mathcal{N}\left(\{q^{\hat{\rho}}_{i_1,\ldots,i_k}\}\right)$, then $\hat{\rho} = \sum_n p_n |\Psi_n\rangle \langle \Psi_n|$, where each $|\Psi_n\rangle$ maximizes $\mathcal{N}\left(\{q^{|\Psi_n\rangle\langle\Psi_n|}_{i_1,\ldots,i_k}\}\right)$. By the triangle inequality, $|\langle a^{(k)}_{i_k}|\hat{\rho}|a^{(1)}_{i_1}\rangle| \leq \sum_n p_n |\langle a^{(k)}_{i_k}|\Psi_n\rangle \langle \Psi_n|a^{(1)}_{i_1}\rangle|$, with equality only if $\arg\left(\langle a^{(k)}_{i_k}|\Psi_n\rangle \langle \Psi_n|a^{(1)}_{i_1}\rangle\right)$ is independent of $n$. So, if $\hat{\rho}$ maximizes $\mathcal{N}\left(\{q^{\hat{\rho}}_{i_1,\ldots,i_k}\}\right)$, then, for each $i_1$ and $i_k$, $\arg\left(\langle a^{(k)}_{i_k}|\Psi_n\rangle \langle \Psi_n|a^{(1)}_{i_1}\rangle\right)$ is independent of $n$. Thus, since $|\langle a^{(k)}_{i_k}|\Psi_n\rangle \langle \Psi_n|a^{(1)}_{i_1}\rangle| = \frac{1}{d}$ for all $n$, $i_1$, and $i_k$, $\langle a^{(k)}_{i_k}|\Psi_n\rangle \langle \Psi_n|a^{(1)}_{i_1}\rangle$ is independent of $n$ for every $i_1, i_k$. Therefore, $|\Psi_n\rangle \langle \Psi_n|$ is independent of $n$, and so $\rho$ is a pure state, as claimed. □

## Appendix F: Real MUBs used to maximize $\mathcal{N}^{\Re^-}$

A Kirkwood-Dirac distribution achieves its maximal negativity when $\mathcal{N}^{\Re^-} = \max\{\mathcal{N}\}$. Such a distribution can be constructed from a triplet of real MUBs. For our purposes, a real MUB is an MUB whose vectors can be represented, relative to some basis, as columns of real numbers. We now reconcile that definition with the definition in the literature.

Real MUBs have been defined as MUBs for Hilbert spaces over $\mathbb{R}^m$, for $m = 2, 3, \ldots$ [72]. In contrast, we focus on Hilbert spaces over $\mathbb{C}^m$. But real MUBs can be imported into complex vector spaces, as follows.

Let $\{B_1, B_2, \ldots, B_n\}$ denote a set of real MUBs for $\mathbb{R}^m$, and let $B_j = \{|b^{(j)}_1\rangle, \ldots, |b^{(j)}_m\rangle\}$. Each vector in $\mathbb{R}^m$ exists in $\mathbb{C}^m$, so each $|b^{(j)}_k\rangle$ exists in $\mathbb{C}^m$. Consider any vector $|\mathrm{v}\rangle$ that exists in $\mathbb{C}^m$ but not in $\mathbb{R}^m$. $|\mathrm{v}\rangle$ equals a linear combination, weighted with complex coefficients, of $\mathbb{R}^m$ vectors. Every $\mathbb{R}^m$ vector equals a linear combination of the $|b^{(j)}_k\rangle$. Therefore, $|\mathrm{v}\rangle \in \mathbb{C}^m$ equals a linear combination of the $|b^{(j)}_k\rangle$. So each $B_j$ is a basis for $\mathbb{C}^m$, so $\{B_1, \ldots, B_n\}$ forms a set of MUBs in $\mathbb{C}^m$.

Let $\mathcal{B}$ denote any basis for $\mathbb{R}^m$. Relative to $\mathcal{B}$, every $|b^{(j)}_k\rangle$ can be represented as a column of real numbers, by the definition of $\mathbb{R}^m$. $\mathcal{B}$ forms a basis also for $\mathbb{C}^m$, by the preceding paragraph. Therefore, every $|b^{(j)}_k\rangle$ can be represented, relative to a basis $\mathcal{B}$ for $\mathbb{C}^m$, as a column of real numbers.

---


[1] W. Heisenberg, in *Original Scientific Papers Wissenschaftliche Originalarbeiten* (Springer, 1927) pp. 478–504.
[2] E. H. Kennard, Zeitschrift für Physik **44**, 326 (1927).
[3] L. D. Landau and E. M. Lifshitz, *Quantum mechanics: non-relativistic theory*, Vol. 3 (Elsevier, 2013).
[4] N. Bohr, Nature **121**, 580 (1928).
[5] S. Designolle, M. Farkas, and J. Kaniewski, New Journal of Physics **21**, 113053 (2019).
[6] J. G. Kirkwood, Phys. Rev. **44**, 31 (1933).
[7] P. A. M. Dirac, Rev. Mod. Phys. **17**, 195 (1945).
[8] E. Wigner, Phys. Rev. **40**, 749 (1932).
[9] L. Cohen, Philosophy of Science **33**, 317 (1966).
[10] R. Hudson, Reports on Mathematical Physics **6**, 249 (1974).
[11] M. D. Srinivas and E. Wolf, Phys. Rev. D **11**, 1477 (1975).
[12] J. B. Hartle, Phys. Rev. A **70**, 022104 (2004).
[13] A. E. Allahverdyan, New Journal of Physics **17**, 085005 (2015).
[14] A. N. Kolmogorov, The Mathematical Gazette **35** (1951), 10.2307/3611503.
[15] W. K. Wootters, Annals of Physics **176**, 1 (1987).
[16] H. J. Carmichael, *Statistical methods in quantum optics 1 Master equations and Fokker-Planck equations* (Springer Science & Business Media, 2013).
[17] K. Husimi, Proceedings of the Physico-Mathematical Society of Japan. 3rd Series **22**, 264 (1940).
[18] E. C. G. Sudarshan, Phys. Rev. Lett. **10**, 277 (1963).
[19] R. J. Glauber, Phys. Rev. **131**, 2766 (1963).
[20] L. Mandel and E. Wolf, *Optical coherence and quantum optics* (Cambridge university press, 1995).
[21] A. Einstein, B. Podolsky, and N. Rosen, Phys. Rev. **47**, 777 (1935).
[22] M. Revzen, P. A. Mello, A. Mann, and L. M. Johansen, Phys. Rev. A **71**, 022103 (2005).
[23] R. W. Spekkens, Phys. Rev. Lett. **101**, 020401 (2008).
[24] Y. P. Terletsky, Journ. Exper. Theor. Phys **7**, 1290 (1937).
[25] H. Margenau and R. N. Hill, Progress of Theoretical Physics **26**, 722 (1961).
[26] L. M. Johansen, Physics Letters A **329**, 184 (2004).
[27] L. M. Johansen and A. Luis, Phys. Rev. A **70**, 052115 (2004).
[28] Y. Aharonov, D. Z. Albert, and L. Vaidman, Phys. Rev. Lett. **60**, 1351 (1988).
[29] I. M. Duck, P. M. Stevenson, and E. C. G. Sudarshan, Phys. Rev. D **40**, 2112 (1989).
[30] O. Hosten and P. Kwiat, Science **319**, 787 (2008).
[31] P. B. Dixon, D. J. Starling, A. N. Jordan, and J. C. Howell, Phys. Rev. Lett. **102**, 173601 (2009).
[32] A. M. Steinberg, Phys. Rev. A **52**, 32 (1995).
[33] D. J. Starling, P. B. Dixon, A. N. Jordan, and J. C. Howell, Phys. Rev. A **80**, 041803 (2009).
[34] J. Dressel, M. Malik, F. M. Miatto, A. N. Jordan, and R. W. Boyd, Rev. Mod. Phys. **86**, 307 (2014).
[35] M. F. Pusey, Phys. Rev. Lett. **113**, 200401 (2014).
[36] S. Pang, J. Dressel, and T. A. Brun, Phys. Rev. Lett. **113**, 030401 (2014).
[37] S. Pang and T. A. Brun, Phys. Rev. Lett. **115**, 120401 (2015).
[38] N. Yunger Halpern, B. Swingle, and J. Dressel, Phys. Rev. A **97**, 042105 (2018).
[39] R. Kunjwal, M. Lostaglio, and M. F. Pusey, Phys. Rev. A **100**, 042116 (2019).
[40] R. Jozsa, Phys. Rev. A **76**, 044103 (2007).
[41] H. F. Hofmann, New Journal of Physics **13**, 103009 (2011).
[42] J. Dressel and A. N. Jordan, Phys. Rev. A **85**, 012107 (2012).
[43] J. T. Monroe, N. Y. Halpern, T. Lee, and K. W. Murch, arXiv preprint arXiv:2008.09131 (2020).
[44] L. M. Johansen, Phys. Rev. A **76**, 012119 (2007).
[45] J. S. Lundeen, B. Sutherland, A. Patel, C. Stewart, and C. Bamber, Nature **474**, 188 (2011).
[46] J. S. Lundeen and C. Bamber, Phys. Rev. Lett. **108**, 070402 (2012).
[47] C. Bamber and J. S. Lundeen, Phys. Rev. Lett. **112**, 070405 (2014).
[48] G. S. Thekkadath, L. Giner, Y. Chalich, M. J. Horton, J. Banker, and J. S. Lundeen, Phys. Rev. Lett. **117**, 120401 (2016).
[49] B. Swingle, G. Bentsen, M. Schleier-Smith, and P. Hayden, Phys. Rev. A **94**, 040302 (2016).
[50] K. A. Landsman, C. Figgatt, T. Schuster, N. M. Linke, B. Yoshida, N. Y. Yao, and C. Monroe, Nature **567**, 61 (2019).
[51] N. Y. Halpern, A. Bartolotta, and J. Pollack, Communications Physics **2**, 1 (2019).
[52] J. R. González Alonso, N. Yunger Halpern, and J. Dressel, Phys. Rev. Lett. **122**, 040404 (2019).
[53] R. Mohseninia, J. R. G. Alonso, and J. Dressel, Phys. Rev. A **100**, 062336 (2019).
[54] D. R. M. Arvidsson-Shukur, A. N. O. Gottfries, and C. H. W. Barnes, Phys. Rev. A **96**, 062316 (2017).
[55] D. R. M. Arvidsson-Shukur and C. H. W. Barnes, Phys. Rev. A **99**, 060102 (2019).
[56] D. R. M. Arvidsson-Shukur, N. Yunger Halpern, H. V. Lepage, A. A. Lasek, C. H. W. Barnes, and S. Lloyd, Nature Communications **11**, 3775 (2020).

[57] A. Levy and M. Lostaglio, arXiv preprint arXiv:1909.11116 (2019).
[58] M. Lostaglio, arXiv preprint arXiv:2004.01213 (2020).
[59] R. B. Griffiths, Journal of Statistical Physics **36**, 219 (1984).
[60] S. Goldstein and D. N. Page, Phys. Rev. Lett. **74**, 3715 (1995).
[61] H. F. Hofmann, M. E. Goggin, M. P. Almeida, and M. Barbieri, Phys. Rev. A **86**, 040102 (2012).
[62] H. F. Hofmann, New Journal of Physics **14**, 043031 (2012).
[63] H. F. Hofmann, Phys. Rev. A **89**, 042115 (2014).
[64] H. F. Hofmann, Phys. Rev. A **91**, 062123 (2015).
[65] H. F. Hofmann, The European Physical Journal D **70**, 118 (2016).
[66] J. J. Halliwell, Phys. Rev. A **93**, 022123 (2016).
[67] B. C. Stacey, arXiv preprint arXiv:1907.02432 (2019).
[68] N. Yunger Halpern, Phys. Rev. A **95**, 012120 (2017).
[69] J. Dressel, Phys. Rev. A **91**, 032116 (2015).
[70] N. Lupu-Gladstein *et al.*, "Experimental demonstration of postselection-enhanced metrology and its connection to quantum non-commutation," (in prep).
[71] T. Durt, B.-G. Englert, I. Bengtsson, and K. Życzkowski, International journal of quantum information **8**, 535 (2010).
[72] P. O. Boykin, M. Sitharam, M. Tarifi, and P. Wocjan, arXiv preprint quant-ph/0502024 (2005).