Negative quasiprobabilities enhance phase estimation in quantum-optics experiment

Noah Lupu-Gladstein,[1, *] Y. Batuhan Yilmaz,[1, †] David R. M. Arvidsson-Shukur,[2, 3, ‡] Aharon Brodutch,[1, §] Arthur O. T. Pang,[1, ¶] Aephraim M. Steinberg,[1, **] and Nicole Yunger Halpern[4, 5, 6, 7]

[1]*CQIQC and Department of Physics, University of Toronto,*
60 Saint George St., Toronto, ON M5S 1A7, Canada
[2]*Hitachi Cambridge Laboratory, J. J. Thomson Ave., Cambridge CB3 0HE, United Kingdom*
[3]*Cavendish Laboratory, Department of Physics, University of Cambridge, Cambridge CB3 0HE, United Kingdom*
[4]*Joint Center for Quantum Information and Computer Science,*
NIST and University of Maryland, College Park, MD 20742, USA
[5]*Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742, USA*
[6]*ITAMP, Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA*
[7]*Department of Physics, Harvard University, Cambridge, MA 02138, USA*
(Dated: November 3, 2021)

Operator noncommutation, a hallmark of quantum theory, limits measurement precision, according to uncertainty principles. Wielded correctly, though, noncommutation can boost precision. A recent foundational result relates a metrological advantage with negative quasiprobabilities—quantum extensions of probabilities—engendered by noncommuting operators. We crystallize the relationship in an equation that we prove theoretically and observe experimentally. Our proof-of-principle optical experiment features a filtering technique that we term *partially postselected amplification* (PPA). Using PPA, we measure a waveplate's birefringent phase. PPA amplifies, by over two orders of magnitude, the information obtained about the phase per detected photon. In principle, PPA can boost the information obtained from the average filtered photon by an arbitrarily large factor. The filter's amplification of systematic errors, we find, bounds the theoretically unlimited advantage in practice. PPA can facilitate any phase measurement and mitigates challenges that scale with trial number, such as proportional noise and detector saturation. By quantifying PPA's metrological advantage with quasiprobabilities, we reveal deep connections between quantum foundations and precision measurement.

Introduction.—Advances in quantum metrology have engendered new measurement techniques [1–5]. The paradigmatic quantum measurement is phase estimation, whose applications span polarimetry, magnetic sensing, gravitational-wave astronomy, and quantum-computer calibration [6–12]. A fundamental limit bounds how precisely one can estimate a phase from a given number of trials [13, 14]. If some trials are filtered out, the average information per retained, or *postselected*, trial can exceed this limit [15]. Filtering can never increase the information per *input* trial, so successful postselections' rarity counterbalances the extra information [16, 17]. Nevertheless, distilling information from many input trials into fewer postselected trials can alleviate challenges that scale with trial number, including detector saturation, proportional noise, low-frequency noise, limited memory, and limited computational power [18–22].

We elucidate this distillation's physical and mathematical roots using a filtering technique that we call *partially postselected amplification* (PPA). Theoretically, the information obtained per PPA trial can diverge as the fraction of postselected trials vanishes [15]. A related technique, weak-value amplification, offers a similarly diverging advantage [18, 20–36]. Both techniques are examples of *noncommutative filtering.* We define *noncommutative filtering* as any filtering whose effect does not depend on when the filter acts. During the alternative, *commutative filtering*, the per–postselected-trial precision cannot exceed the per–input-trial limit [15]. Examples include the neutral-density filter that reduces a camera's overexposure. PPA's postselected trials break the per–input-trial limit by endowing a certain quasiprobability distribution with negative elements [15].

Quasiprobabilities represent quantum states as probability densities represent states in classical statistical mechanics. Like probabilities, the quasiprobabilities in a distribution sum to one. Yet quasiprobabilities can assume negative and nonreal values, called *nonclassical values*. They can arise when the quasiprobability describes quantum-incompatible operations or observables. Well-known quasiprobability distributions include the Wigner function. A rising star is the *Kirkwood-Dirac distribution* [37, 38], which has recently found applications in quantum state tomography [39–43], chaos [44–48], postselected metrology [15, 27, 28, 30, 49–55], measurement disturbance [56–59], quantum thermodynamics [44, 60–63], and quantum foundations [36, 57, 64–75]. Negative Kirkwood-Dirac quasiprobabilities have been demonstrated, under certain conditions, to underlie operational advantages in quantum computation, work extraction, and parameter estimation [15, 46, 55, 63].

---

* nlupugla@physics.utoronto.ca. The first two coauthors contributed equally.
† ybylmaz@physics.utoronto.ca.
‡ drma2@cam.ac.uk
§ brodutch@physics.utoronto.ca
¶ arthur.pang@mail.utoronto.ca
** steinberg@physics.utoronto.ca

$\tilde{p}_{\rho(\theta),t}(a, a'\,\vert+)$	$a' = a_+$	$a' = a_-$
$a = a_+$	$\dfrac{1+\vert t\vert^2}{4p^{\mathrm{ps}}(\theta,t)}$	$e^{i\Delta\theta}\,\dfrac{-1+\vert t\vert^2}{4p^{\mathrm{ps}}(\theta,t)}$
$a = a_-$	$e^{-i\Delta\theta}\,\dfrac{-1+\vert t\vert^2}{4p^{\mathrm{ps}}(\theta,t)}$	$\dfrac{1+\vert t\vert^2}{4p^{\mathrm{ps}}(\theta,t)}$

TABLE I: Conditional Kirkwood-Dirac distribution (7) for our PPA experiment and $\rho(\theta) = \vert\Psi(\theta)\rangle\langle\Psi(\theta)\vert$.

We demonstrate PPA's parameter-estimation enhancement in a proof-of-principle polarimetry experiment. We estimate the birefringent phase imparted to photons by a near-half–waveplate. A tunable polarization filter implements the PPA. The filter boosts the per–detected-photon precision by over two orders of magnitude. Furthermore, we measure a Kirkwood-Dirac distribution that describes the experiment. Our experiment operationally motivates a measure of the distribution's negativity. We prove theoretically and confirm experimentally that the negativity is proportional to the precision enhancement when the phase is probed optimally. We also pinpoint which systematic errors limit PPA's theoretically unbounded precision enhancement (App. A). Our experiment unifies theoretical quantum foundations with practical precision measurement.

Theoretical background and equality.—Consider estimating a parameter θ by measuring a quantum state $\rho(\theta)$. The *quantum Fisher information* (QFI) $\mathcal{I}(\theta)$ quantifies the information provided by $\rho(\theta)$ about θ, via the state's sensitivity to changes in θ [76] (App. B). The QFI's reciprocal lower-bounds the variance of every unbiased estimator θ_{e} of θ, in the Cramér-Rao bound, $\mathrm{var}(\theta_{\mathrm{e}}) \geq 1/\mathcal{I}(\theta)$. [13, 14]

Let A denote an observable with greatest and least eigenvalues a_+ and $a_- = a_+ - \Delta$. The eigenstates $\vert a_\pm\rangle$ satisfy $A\vert a_\pm\rangle = a_\pm\vert a_\pm\rangle$. Let a unitary $U(\theta) = \exp(i\theta A)$ imprint θ on an input state. The optimal inputs are even-weight superpositions of extremal A eigenstates, e.g., $\vert 0\rangle = (\vert a_+\rangle + \vert a_-\rangle)/\sqrt{2}$ and $\vert 1\rangle = (\vert a_+\rangle - \vert a_-\rangle)/\sqrt{2}$. The imprinted state $U(\theta)\vert 0\rangle = \vert\Psi(\theta)\rangle$ carries the most QFI possible without postselection, $\mathcal{I}(\theta) = \Delta^2$.

A postselected state can provide more QFI. If the angle is small ($\theta\Delta \ll 1$), then $\vert\Psi(\theta)\rangle \approx \vert 0\rangle + i\frac{\theta\Delta}{2}\vert 1\rangle$. The $\vert 0\rangle$ coefficient is less sensitive to θ than the $\vert 1\rangle$ coefficient, yet $\vert 0\rangle$ has a greater population. PPA partially postselects on $\vert 1\rangle$ via a filter whose $\vert 1\rangle$ transmission amplitude is unity and whose $\vert 0\rangle$ transmission amplitude is parametrically smaller.

More precisely, let t denote the amplitude for $\vert 0\rangle$'s survival of the filter. The filter acts as the Kraus operator [77] $K(t) = t\vert 0\rangle\langle 0\vert + \vert 1\rangle\langle 1\vert$, wherein $\vert t\vert \in [0, 1]$. The filter lets $\vert\Psi(\theta)\rangle$ pass with a probability

$$p^{\mathrm{ps}}(\theta, t) = \mathrm{Tr}\big(K(t)\vert\Psi(\theta)\rangle\langle\Psi(\theta)\vert K(t)^\dagger\big) \quad (1)$$

$$= \vert t\vert^2\cos^2(\Delta\theta/2) + \sin^2(\Delta\theta/2). \quad (2)$$

The state becomes

$$\vert\Psi^{\mathrm{ps}}(\theta, t)\rangle = K(t)\vert\Psi(\theta)\rangle/\sqrt{p^{\mathrm{ps}}(\theta, t)} \quad (3)$$

$$= \cos(\Delta\Theta/2)\vert 0\rangle + i\sin(\Delta\Theta/2)\vert 1\rangle. \quad (4)$$

The filter effectively amplifies θ to a Θ defined through $\tan(\Delta\Theta/2) = \tan(\Delta\theta/2)/\vert t\vert$. The postselected state carries the QFI

$$\mathcal{I}(\theta) = [\Delta\,\vert t\vert/p^{\mathrm{ps}}(\theta, t)]^2. \quad (5)$$

A large angle is typically easier to observe than a smaller one. If the angle is small, $\Delta\theta \ll 1$, then Θ exceeds θ by a factor $1/\vert t\vert$. This amplification boosts the information obtained per detected state: $\mathcal{I}(\theta) \approx (\Delta/\vert t\vert)^2$. The amplification is arbitrarily large if $\Delta\theta$ is arbitrarily small. Such extreme filtering does not significantly reduce the information obtainable per input state: $p^{\mathrm{ps}}(\theta, t)\mathcal{I}(\theta) \approx \Delta^2$, if $\tan(\Delta\theta/2) \ll \vert t\vert$.

PPA can be beneficial even if $\Delta\theta$ is large. Suppose prior knowledge indicates that $\theta \approx \theta_{\mathrm{p}} \in [0, 2\pi)$. Performing $U(-\theta_{\mathrm{p}})$ after $U(\theta)$ shrinks the probed angle to $\Delta(\theta - \theta_{\mathrm{p}})$.

Why can a successful PPA trial offer more information than Δ^2, the most information offered by any input trial? Reference [15] identified a necessary condition. A projectively postselected trial can carry information $> \Delta^2$ only if a Kirkwood-Dirac distribution contains a negative quasiprobability. We generalize that result beyond projective postselection.

Let $\{\vert a\rangle\}_a$ and $\{\vert a'\rangle\}_{a'}$ denote copies of an A eigenbasis. Kraus operators $\{K_f\}_f$ with $\sum_f K_f^\dagger K_f = \mathbb{1}$ model the partial postselection. The information-bearing state $\rho(\theta)$ is represented by the Kirkwood-Dirac quasiprobabilities (App. C)

$$\tilde{p}_{\rho(\theta)}(a, f, a') := \mathrm{Tr}\big(\vert a'\rangle\langle a'\vert K_f^\dagger K_f\vert a\rangle\langle a\vert\rho(\theta)\big). \quad (6)$$

Conditioning on a postselection outcome f induces the *conditional Kirkwood-Dirac distribution*

$$\tilde{p}_{\rho(\theta)}(a, a'\vert f) := \tilde{p}_{\rho(\theta)}(a, f, a')/\sum_{a,a'}\tilde{p}_{\rho(\theta)}(a, f, a'). \quad (7)$$

These quasiprobabilities are positive if A and $K_f^\dagger K_f$ commute on the support of $\rho(\theta)$ [73].

PPA involves Kraus operators, $K_+ = K(t)$ and $K_- = \sqrt{\mathbb{1} - K(t)^\dagger K(t)}$, that effect successful and unsuccessful postselection. Table I shows PPA's conditional quasiprobabilities, labeled by t, for $\rho(\theta) = \vert\Psi(\theta)\rangle\langle\Psi(\theta)\vert$. If $\Delta\theta < \pi$ and $\vert t\vert^2 < 1$, the real part of $\tilde{p}_{\rho(\theta),t}(a_\pm, a_\mp\vert+)$ is negative, and the postselected QFI (5) exceeds Δ^2. This concurrence stems from an equality that we prove.

Let x denote the vector of arguments for a Kirkwood-Dirac distribution $\{\tilde{p}(x)\}_x$. Define the *nonclassicality gap* as the greatest difference between quasiprobabilities' absolute squares:[1] $\max_x\{\vert\tilde{p}(x)\vert^2\} - \min_x\{\vert\tilde{p}(x)\vert^2\}$. The gap > 1 only if a quasiprobability $\notin [0, 1]$. For any

---

[1] See [46, 73, 74] for alternative negativity measures. Equation (8) motivates our definition.
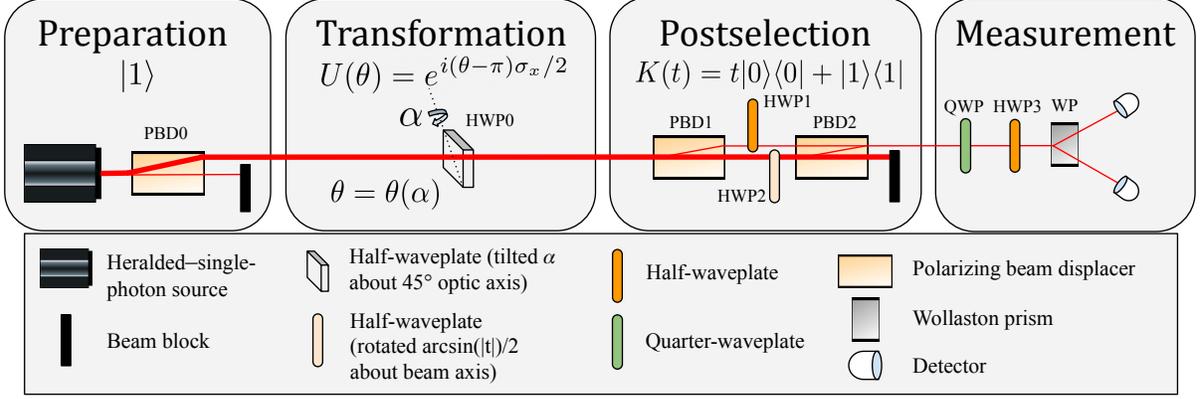
FIG. 1: Photonic parameter-estimation experiment: Preparation: A heralded–single-photon source emits vertically polarized ($|1\rangle$) light. Transformation: The half-waveplate (HWP0) has an optic axis angled 45° above the horizontal. HWP0 is tilted away from normal incidence through an angle α about its optic axis. The waveplate rotates a photon's polarization through an angle $\theta(\alpha) - \pi$. A calibration curve of $\theta(\alpha) \equiv \theta$ provides a prior estimate of θ. We use this estimate to calculate the polarization projection optimal for inferring θ (App. B). Postselection: A polarizing–beam-displacer interferometer, followed by a beam block in the undisplaced port, realizes a partial polarizer. The horizontal-polarization transmission amplitude, t with $|t| \in [0, 1]$, is controlled by a half-waveplate (HWP2) inside the interferometer. The filter discards all horizontally polarized photons when $|t| = 0$ and none when $|t| = 1$. Measurement: Motorized waveplates, followed by a Wollaston prism (WP) and single-photon counters, project onto any desired polarization.


(a)   **(b)**   **(c)**
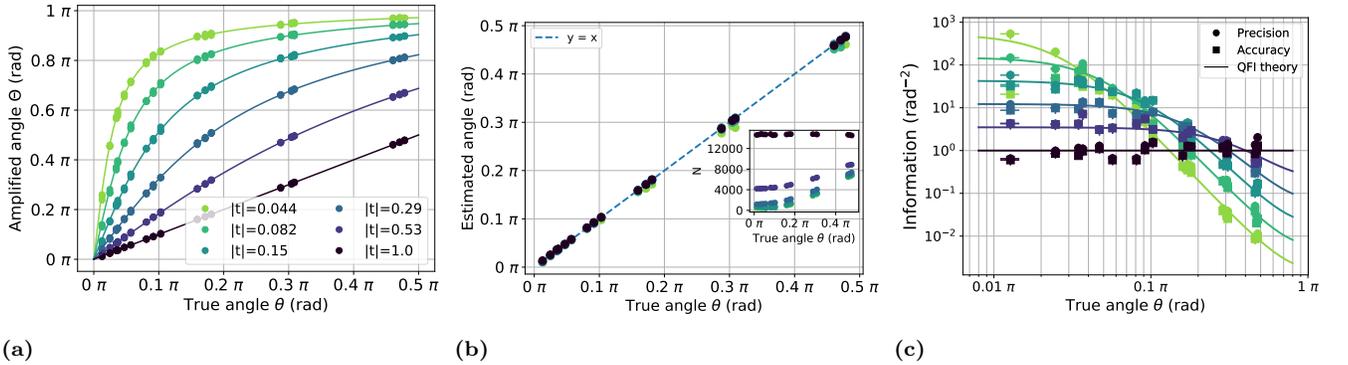
FIG. 2: Experimental performance of PPA. Color represents the postselection-parameter magnitude $|t|$. (a) Amplified angle vs. true angle. The slope signifies sensitivity to changes in θ. The data adhere to the prediction (solid lines), $\tan(\Delta\Theta/2) = \tan(\Delta\theta/2)/|t|$. When θ is small [$\tan(\Delta\theta/2) \ll |t|$], PPA magnifies θ by a factor $1/|t|$. Setting $|t| = \tan(\Delta\theta/2)$ amplifies θ to $\pi/2$ and optimizes the sensitivity. Decreasing $|t|$ further reduces the sensitivity, rendering prior knowledge about θ important. (b) Mean estimate of θ vs. true θ value. Each mean is averaged over 32 independent estimates and agrees roughly with the true value (App. A details systematic errors). Inset: mean detected-photon number, N. (c) Information per photon vs. θ. Solid lines represent the predicted QFI (5). Circles represent the 32 estimates' precision (1/variance) per mean detected photon. Squares show accuracy per detected photon (1/[mean squared error] $\equiv$ 1/MSE). All error bars follow from assuming that the variances follow a scaled χ^2 distribution. The per-photon precision agrees excellently with theory and climbs to 540 ± 150 rad^{-2} at $(\theta, |t|) = (0.040 \text{ rad}, 0.044)$. The per-photon accuracy suffers from systematic errors at the smallest θ and $|t|$, yet still reaches 78 ± 15 rad^{-2} at $(\theta, |t|) = (0.116 \text{ rad}, 0.082)$.

postselection operator K_+, the nonclassicality gap is proportional to the optimal input state's postselected QFI (App. D):

$$\mathcal{I}(\theta) = 4\,\Delta^2$$
$$\times \left[\max_{a,a'}\left\{|\tilde{p}_{\rho(\theta)}(a, a'|+)|^2\right\} - \min_{a,a'}\left\{|\tilde{p}_{\rho(\theta)}(a, a'|+)|^2\right\}\right]. \tag{8}$$

Equation (8) crystallizes the relationship between postselected quantum metrology and Kirkwood-Dirac nonclassicality.

Experimental setup.—We realize PPA in a proof-of-principle polarimetry experiment (Fig. 1). The to-be-estimated parameter θ is the excess birefringent phase, beyond π, imparted by a near-half–waveplate (HWP0).

(a)   **(b)**
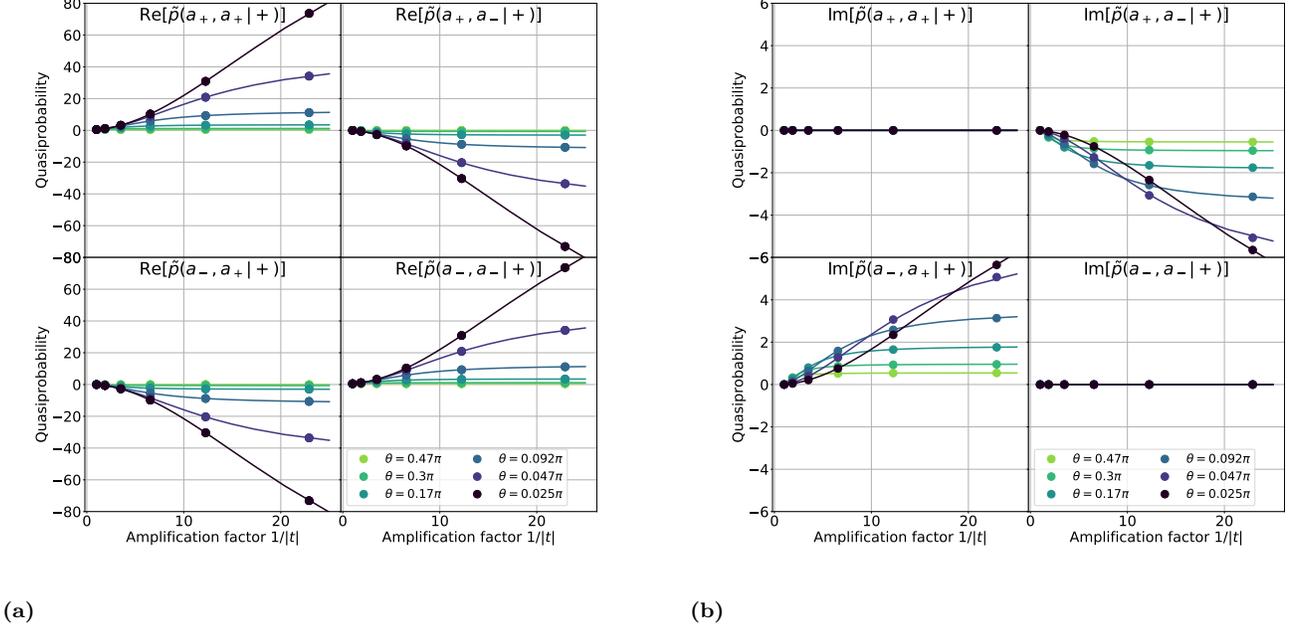
FIG. 3: Quasiprobabilities vs. amplification factor $1/|t|$. We inferred the Kirkwood-Dirac distribution (6), $\tilde{p}_{\rho(\theta),t}(a,a'|+)$, from state tomography. The arguments $a_\pm = \pm 1/2$. (a) Real component of distribution. (b) Imaginary component. At each θ and $|t|$, the quasiprobabilities are complex, yet sum to unity. As the amplification strengthens, the components' magnitudes increase. The off-diagonal quasiprobabilities' real components become extremely negative: < -70 at the smallest θ and $|t|$ values.

A heralded–single-photon source emits vertically polarized photons with wavelengths of 808 nm. The photons hit HWP0, whose optic axis lies 45° above the horizontal. Tilting HWP0 through an incidence angle α sets its birefringent retardance to $\theta(\alpha) - \pi$. A calibration curve of $\theta(\alpha) \equiv \theta$ provides prior knowledge about θ.

Denote horizontal polarization by $|0\rangle$; and vertical polarization, by $|1\rangle$. We filter the photons by attenuating one polarization, using an interferometer formed from polarizing beam displacers. The postselection parameter t equals (the filter's $|0\rangle$ transmission amplitude)/(the $|1\rangle$ transmission amplitude). We control t with a motorized waveplate (HWP2) placed in the interferometer.

HWP0 rotates the photon's polarization with the unitary $\exp(i[\theta - \pi]\sigma_x/2)$. The generator $A = \sigma_x/2$ has eigenvalues $a_\pm = \pm 1/2$ and eigenstates $|a_\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$. The filtered photons occupy the state $\rho^{\mathrm{ps}}(\theta,t)$—ideally, the pure state (4). We projectively measure the state's polarization to estimate θ.

Experimental results.—First, we assess PPA's metrological performance. Then, we present the measured quasiprobabilities (7). Comparing the quasiprobabilities with the QFI, we support Eq. (8) experimentally.

Polarization tomography reveals how PPA boosts sensitivity. Figure 2a shows the postselected state's amplified angle, Θ, versus the true θ value. We infer the latter using state tomography without postselection ($|t| = 1$).

The slope of $\Theta(\theta)$ quantifies our sensitivity to small changes in θ. When $|t| = 1$, $\Theta(\theta)$ has a unit slope. As we postselect more ($|t|$ decreases), the slope grows—by a factor of > 20 at $|t| = 0.044$.

We estimate θ by projectively measuring many copies of the amplified state identically. The measurement basis, which we optimize according to calibrations of $\theta(\alpha)$ and t, provides the QFI (App. B). For each (θ, t), we sample 32 independent estimates of θ. Figure 2b displays the mean estimate versus θ. The mean estimates roughly agree with the true values. Appendix A details systematic errors.

Figure 2c displays our estimates' precision and accuracy, normalized by the number N of detected photons. The precision per photon, $\mathrm{var}(\theta_\mathrm{e})^{-1}/N$, agrees excellently with the QFI (5). The accuracy per photon, $\mathrm{MSE}(\theta_\mathrm{e})^{-1}/N$, mostly agrees with the QFI but falls short at the smallest θ and $|t|$. The per-photon precision enhancement maximizes at 540 ± 150, when $\theta = 0.040$ rad, $|t| = 0.044$. The per-photon accuracy caps at 78 ± 15, when $\theta = 0.116$ rad and $|t| = 0.082$. The discrepancy between precision and accuracy arises because PPA amplifies systematic errors (App A).

We extract the conditional quasiprobabilities (6) from tomography of the unpostselected ($|t| = 1$) state. Figure 3 shows the quasiprobabilities' real and imaginary parts. At each (θ, t), the quasiprobabilities sum to one. When
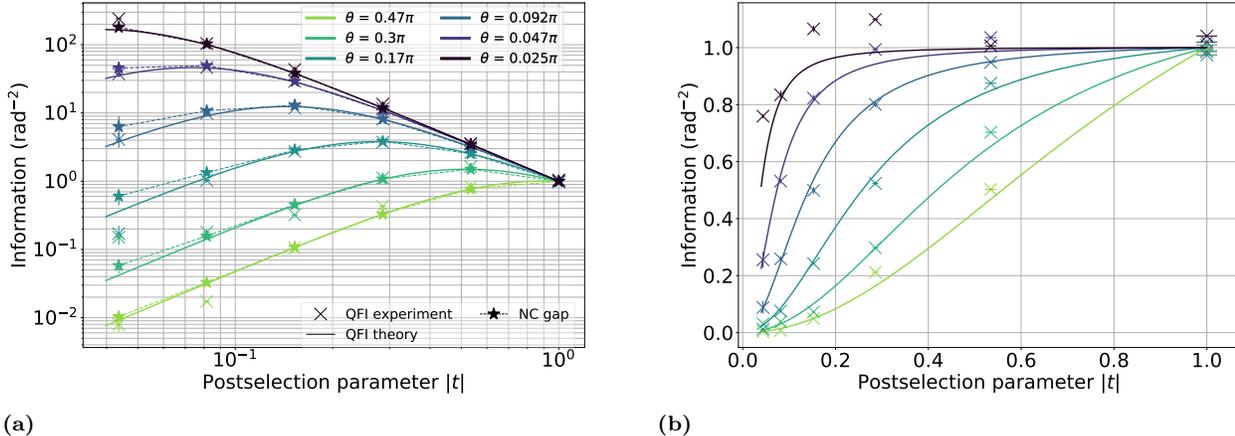
FIG. 4: Information per detected photon (a) and information per input photon (b) vs. postselection-parameter magnitude $|t|$. Each marker shape signifies a metrological quantity. Vertical error bars denote the standard error of 4 independent runs. In (a), each quantity is divided by the number of detected photons. In (b), each quantity is divided by the number of photons detected without postselection ($|t| = 1$). Crosses represent the QFI estimated from state tomography. Solid lines represent the theoretical QFI prediction (5). Dashed lines with stars represent 4 times the empirical nonclassicality gap. Without postselection, our estimates are shot-noise–limited to the per–input-photon precision 1 rad^{-2}. As we increasingly postselect (as $|t|$ decreases), the per–detected-photon precision increases when $\theta \approx 0$ and decreases when $\tan(\theta/2) < |t|$, agreeing with theory. The smallest $|t|$ and θ provide a per–detected-photon precision > 200 rad^{-2}, despite sacrificing little per–input-photon precision.

$|t| < 1$, quasiprobabilities acquire negative real parts, so other quasiprobabilities acquire real parts > 1 to ensure a unit sum. As $|t|$ decreases, elements' magnitudes increase—to > 70 at the smallest θ and $|t|$.

Figure 4 compares the nonclassicality gap with the QFI. We compute the gap from the quasiprobabilities shown in Fig. 3. We determine the QFI at $\theta = \theta_0$ empirically from estimates of $\rho^{\mathrm{ps}}(\theta_0, t)$ and $\partial \rho^{\mathrm{ps}}(\theta, t)/\partial \theta|_{\theta_0}$. The derivative is the matrix slope of a linear fit through three tomographic estimates: $\rho^{\mathrm{ps}}(\theta_0 - d\theta, t), \rho^{\mathrm{ps}}(\theta_0, t)$, and $\rho^{\mathrm{ps}}(\theta_0 + d\theta, t)$; $d\theta = 0.035$ radians. We estimate the nonclassicality gap's and QFI's uncertainties by performing tomography four times and assuming Poissonian photon statistics. The empirical QFI and nonclassicality gap are consistent with the theoretical QFI [Fig. 4(a)]. Thus, our experiment corroborates the relationship (8) between enhanced precision and quasiprobability negativity.

Conclusions.—We have experimentally demonstrated and theoretically proved how negative Kirkwood-Dirac quasiprobabilities enhance postselected metrology. We introduced and illustrated a scheme for phase estimation, partially postselected amplification. In our polarimetry experiment, PPA boosted our per–detected-photon precision by over two orders of magnitude. This enhancement derives from negativity of a generalized Kirkwood-Dirac quasiprobability, according to an equation that we prove and experimentally support. The negativity demonstrates that our filter provides a benefit offered by no filter that commutes with $U(\theta)$.

In theory, PPA's precision boost is unbounded. In practice, we find, the phase amplification augments systematic errors. Yet the error amplification has a silver lining, having helped us detect and correct such errors in our implementation of the generator A (App. A).

PPA is related to weak-value amplification (WVA), a scheme for estimating couplings strengths [18, 20, 23–36]. PPA and WVA concentrate information spread across many input trials into few postselected trials. Yet PPA differs from WVA in three ways: (i) PPA can amplify any phase, not just coupling strengths. (ii) PPA survives decoherence better. In WVA, an interaction couples two systems. One system is measured, the other is postselected, and both must remain coherent during the interaction. PPA only requires the measured system to maintain coherence. (iii) PPA admits of a simpler mathematical treatment: WVA requires a Hilbert-space dimensionality ≥ 4, whereas PPA works with a Hilbert-space dimensionality ≥ 2. PPA is therefore a promising tool for combating metrological challenges that scale with the number of completed trials. As a whole, our work interweaves the disparate studies of precision measurement and quantum foundations.
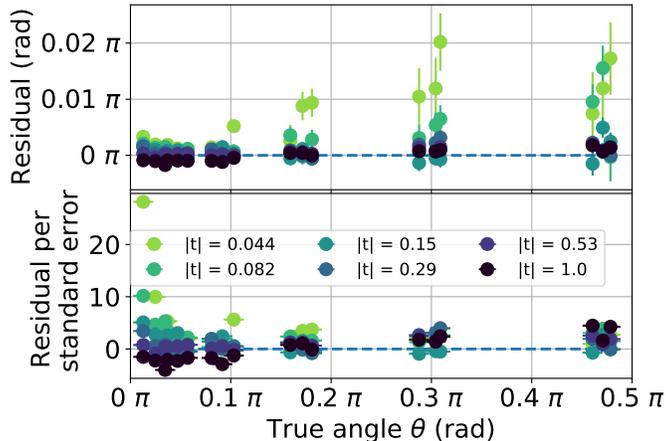
Appendix A   POSTSELECTED AMPLIFICATION OF SYSTEMATIC ERROR

FIG. 5: Difference between our estimates and the true θ value. In the lower graph, the difference is normalized by standard error. Modest postselections ($|t| \geq 0.15$) do not suffer from significantly more systematic error than postselection-free measurements ($|t| = 1$). The two most severe postselections ($|t| = 0.044, 0.082$) amplify systematic errors significantly.

Some estimates of θ (Fig. 2b) deviate statistically significantly from the true θ values. We now detail these deviations and explain PPA's ability to amplify certain types of systematic errors.

Figure 5 shows our mean estimates' residual, and residual per standard error, versus the true θ value. Deviations greater than 2 standard errors point to statistically significant systematic errors. Figure 2b showed rough agreement between our estimates and true values. However, Fig. 5 reveals that statistically significant systematic errors bedevil our estimates.

We now distinguish the systematic errors present with and without postselection. Without postselection ($|t| = 1$), our estimates of θ differ significantly from the true (tomographically extracted) values. This error is not unique to PPA and is specific to our polarimetry experiment's details. Another type of error emerges at the two most extreme postselection levels, when $|t| \in \{0.044, 0.082\}$. The systematic errors varies strongly with θ and characterizes PPA.

While amplifying θ, PPA also amplifies small errors in the implementations of (i) the generator A and (ii) the Kraus operator K. As shown in the main text, PPA amplifies an unknown phase θ to a larger phase Θ. To obtain an estimate θ_{e} of θ, we first obtain an estimate Θ_{e} of Θ. Then, we numerically invert the amplification. Inversion requires an accurate model of the experiment, but no model is perfect. In our model, postselection amplifies the state's phase to

$$\Theta = \Theta(\theta, t) = 2\arctan\big(\tan(\theta/2)/t\big), \tag{A1}$$

according to Eq. (4). Given the amplified-angle estimate, the θ estimate is $\theta_{\mathrm{e}} = \Theta^{-1}(\Theta_{\mathrm{e}}, t)$.

The setting of t may suffer from a small error Δt. The estimated phase, as a function of the true phase, will become

$$\theta_{\mathrm{e}} = 2\arctan\big(\tan(\theta/2)[1 + \Delta t/t]\big). \tag{A2}$$

The half-tangents of θ_{e} and θ differ by

$$\tan(\theta_{\mathrm{e}}/2) - \tan(\theta/2) = [\tan(\theta/2)/t]\,\Delta t = \tan(\Theta/2)\,\Delta t. \tag{A3}$$

This difference is proportional to the amplified half-tangent, which leads to two problems. First, amplifying θ amplifies the effects of the uncertainty in t. Second, the resulting systematic error cannot be corrected, because $\tan(\Theta/2)$ is unknown, by the phase-estimation task's nature.

Separately, errors can mar the implementation of the generator A. Ideally, $A = \sigma_x/2$. However, A is effected by a waveplate whose optic axis may be rotated by a small angle ϵ from $45°$ to the horizontal. The generator will become $A(\epsilon) = \cos(2\epsilon)\sigma_x/2 + \sin(2\epsilon)\sigma_z/2$. Under $e^{i(\theta - \pi)A(\epsilon)}$, the initial state $|1\rangle$ evolves to

$$|\psi(\epsilon)\rangle = \cos(\theta/2)[\cos(2\epsilon)|0\rangle - \sin(2\epsilon)|1\rangle] + i\sin(\theta/2)|1\rangle. \tag{A4}$$

This state's polar half-angle has a tangent

$$\frac{|\langle 1|\psi(\epsilon)\rangle|}{|\langle 0|\psi(\epsilon)\rangle|} = \sqrt{\frac{\sin^2(2\epsilon) + \tan^2(\theta/2)}{\cos^2(2\epsilon)}} \,. \tag{A5}$$

When $\sin^2(2\epsilon) \ll \tan^2(\theta/2)$, ϵ affects the polar angle negligibly. However, if $\sin^2(2\epsilon) > \tan^2(\theta/2)$, the polar angle depends on ϵ considerably. Amplifying the polar angle with PPA amplifies also the error in the A implementation.

One can compensate, or correct, for the error amplification. Moreover, we exploited it to improve our A implementation while aligning our equipment. However, calibration and correction errors propagate to θ eventually. Thus, the precision of one's $K(t)$ and A implementations limits how much one can filter and postselect usefully. This phenomenon extends beyond our polarimetery experiment to every PPA implementation.

Appendix B   OPTIMAL MEASUREMENT

We estimated θ with an optimal polarization measurement. By *optimal measurement*, we mean a measurement whose information yield, averaged over many trials, equals the QFI. Let Λ denote the symmetric logarithmic derivative of ρ, defined implicitly through the Sylvester equation,

$$\frac{\partial \rho}{\partial \theta} = \frac{\Lambda \rho + \rho \Lambda}{2}. \tag{B1}$$

Projectively measuring an eigenbasis of Λ is optimal [76]. Therefore, we aim to solve Eq. (B1) for Λ.

We obtain a solution as follows. Let $\text{vec}(Y)$ denote the vectorized form of a matrix Y in row-major order. For matrices X, Y, and Z with appropriate dimensions, $\text{vec}(XYZ) = (X \otimes Z^T)\text{vec}(Y)$. Let Y^+ denote the Moore-Penrose inverse of a matrix Y. Solving for Λ, we vectorize it:

$$\text{vec}(\Lambda) = \left(\frac{\rho \otimes \mathbb{1} + \mathbb{1} \otimes \rho^T}{2}\right)^+ \text{vec}\left(\frac{\partial \rho}{\partial \theta}\right). \tag{B2}$$

In our experiment, the photon polarization is nearly, but not exactly, pure: The initial polarization is $\rho_0 = v|0\rangle\langle 0| + (1-v)\mathbb{1}/2$. The visibility v is close, but not equal, to unity. The postselected state is

$$\rho^{\text{ps}}(\theta, t) = K(t)U(\theta)\rho_0 U(\theta)^\dagger K(t)^\dagger / p^{\text{ps}}(\theta, t), \tag{B3}$$

and the postselection probability is

$$p^{\text{ps}}(\theta, t) = \text{Tr}\big(K(t)U(\theta)\rho_0 U(\theta)^\dagger K(t)^\dagger\big). \tag{B4}$$

Λ assumes the form

$$\Lambda = v\, p^{\text{ps}}(\theta, t)\left[\frac{1 - |t|^2}{2}\sin(\theta)\mathbb{1} + \cos(\theta)(Re[t]\sigma_x + Im[t]\sigma_y) + \frac{1 + |t|^2}{2}\sin(\theta)\sigma_z\right]. \tag{B5}$$

An optimal measurement is thus a polarization onto the Bloch vector whose polar angle θ_{opt} and azimuthal angle ϕ_{opt} are defined through

$$\cot(\theta_{\text{opt}}) = \frac{1 + |t|^2}{2|t|}\tan(\theta), \quad \tan(\phi_{\text{opt}}) = \Im[t]/\Re[t]. \tag{B6}$$

θ_{opt} turns out to be independent of v but not of θ. We use prior knowledge about θ, derived from our calibration curve $\theta(\alpha)$, to estimate θ_{opt} and approximate an optimal measurement.

Appendix C   GENERALIZED KIRKWOOD-DIRAC DISTRIBUTION

We called the quasiprobabilities (6) "Kirkwood-Dirac quasiprobabilities." This appendix justifies that terminology. Kirkwood and Dirac independently defined a quasiprobability dependent on a pure state $|\psi\rangle$ and on two projectors, $|a\rangle\langle a|$ and $|f\rangle\langle f|$ [37, 38]: $\text{Tr}(|f\rangle\langle f|a\rangle\langle a|\psi\rangle\langle\psi|)$. Wiseman generalized the pure state to an arbitrary quantum state ρ [78]: $\text{Tr}(|f\rangle\langle f|a\rangle\langle a|\rho)$. Yunger Halpern and collaborators generalized to arbitrarily many projectors $\Pi_{m_j}^{M_j}$, which can

project onto multidimensional subspaces [44, 45]: $\mathrm{Tr}\left(\Pi_{m_k}^{(k)}\Pi_{m_{k-1}}^{(k-1)}\ldots\Pi_{m_1}^{(1)}\rho\right)$. They called this quantity an *extended Kirkwood-Dirac quasiprobability*.

We generalize the projectors to elements of POVMs $\left\{M_{m_1}^{(1)}\right\}_{m_1},\left\{M_{m_2}^{(2)}\right\}_{m_2},\ldots,\left\{M_{m_k}^{(k)}\right\}_{m_k}$:

$$\tilde{p}_\rho(m_1, m_2, \ldots, m_k) := \mathrm{Tr}\left(M_{m_k}^{(k)}M_{m_{k-1}}^{(k-1)}\ldots M_{m_1}^{(1)}\rho\right). \tag{C1}$$

By definition, the ℓth POVM's elements are positive-semidefinite operators that resolve unity: $M_{m_\ell}^{(\ell)} \geq 0$, and $\sum_{m_\ell} M_{m_\ell} = \mathbb{1}$. We could call our quasiprobability (C1) a "generalized extended Kirkwood-Dirac quasiprobability"; but the name, itself, would be too extended. Following Wiseman's lead, we avoid new names; we christen our construction, and rechristen the extended Kirkwood-Dirac quasiprobability, a "Kirkwood-Dirac quasiprobability."

To merit the name, our quasiprobability distribution should exhibit two recursive properties:

1. A Kirkwood-Dirac distribution with one argument is a probability distribution (albeit a trivial one).

2. Marginalizing over any argument of a Kirkwood-Dirac distribution with $k > 1$ arguments yields a Kirkwood-Dirac distribution with $k - 1$ arguments.

Consequently, marginalizing over $k - 1$ arguments of a k-argument Kirkwood-Dirac distribution yields a probability distribution. We now verify that our distribution (C1) has these properties.

Consider the one-argument distribution with elements $\tilde{p}_\rho(m_1) = \mathrm{Tr}(M_{m_1}^{(1)}\rho)$. It satisfies property 1 because (i) each $M_{m_1}^{(1)}$ is positive-semidefinite, so $\tilde{p}_\rho(m_1) \geq 0$; and (ii) $\left\{M_{m_1}^{(1)}\right\}_{m_1}$ resolves unity, so $\sum_{m_1}\tilde{p}_\rho(m_1) = \mathrm{Tr}(\rho) = 1$. The element $\tilde{p}_\rho(m_1)$ is the probability that, upon preparing ρ and measuring $\left\{M_{m_1}^{(1)}\right\}_{m_1}$, one obtains outcome m_1.

Now, consider the k-argument distribution (C1). Marginalizing over the ℓth POVM yields

$$\sum_{m_\ell}\tilde{p}_\rho(m_1, m_2, \ldots, m_{\ell-1}, m_\ell, m_{\ell+1}, \ldots, m_k) = \tilde{p}_\rho(m_1, m_2, \ldots m_{\ell-1}, m_{\ell+1}, m_{\ell+2}, \ldots, m_k) \tag{C2}$$

$$= \mathrm{Tr}\left(M_{m_k}^{(k)}M_{m_{k-1}}^{(k-1)}\ldots M_{m_{\ell+1}}^{(\ell+1)}M_{m_{\ell-1}}^{(\ell-1)}M_{m_{\ell-2}}^{(\ell-2)}\ldots M_{m_1}^{(1)}\rho\right), \tag{C3}$$

because each POVM resolves unity. The right-hand side is a $(k-1)$-argument Kirkwood-Dirac distribution. Therefore, our construction has property 2.

Appendix D PROOF OF PROPORTIONALITY BETWEEN POSTSELECTED QUANTUM FISHER INFORMATION AND QUASIPROBABILITY NONCLASSICALITY

Equation (8) interrelates the postselected quantum Fisher information and the generalized quasiprobability's nonclassicality. We prove this equation—in fact, a generalization of the equation—here.

Generalized setup: All the definitions presented here are independent of the definitions presented in the main text. For example, the A defined here is independent of the A defined in the main text. However, quantities defined in the main text are examples of quantities defined here.

During a parameter-estimation experiment, the input state ρ undergoes a unitary $U(\theta) = e^{i\theta A}$ generated by a Hermitian operator A:

$$\rho \mapsto U(\theta)\rho U(\theta)^\dagger = \rho(\theta). \tag{D1}$$

The parameter to be estimated is θ. The system corresponds to a Hilbert space of arbitrary dimensionality. We assume that ρ is a pure state with support on only two A eigenspaces. The projectors onto these eigenspaces, we denote by $\Pi_{\min}$ and $\Pi_{\max}$. Their associates eigenvalues are $a_{\min}$ and $a_{\max} > a_{\min}$.

After undergoing the unitary, the state meets a filter. If the state survives the filter, a Kraus operator K_+ updates the state. Otherwise, $K_- = \sqrt{\mathbb{1} - K_+^\dagger K_+}$ updates the state. The state survives the filter with the postselection probability $p^{\mathrm{ps}}(\theta) = \mathrm{Tr}(\rho(\theta) K_+^\dagger K_+)$. We can represent the experiment with the Kirkwood-Dirac quasiprobability

$$\tilde{p}_{\rho(\theta)}(a, f, a') = \mathrm{Tr}(\Pi_{a'} K_f^\dagger K_f \Pi_a \rho(\theta)), \quad \text{for } a, a' \in \{a_{\max}, a_{\min}\}. \tag{D2}$$

Our result depends on the conditional quasiprobabilities, labeled by $f = +$, associated with a successful postselection:

$$\tilde{p}_{\rho(\theta)}(a, a'|+) = \tilde{p}_{\rho(\theta)}(a, +, a')/p^{\text{ps}}(\theta). \tag{D3}$$

We assume that two "diagonal" quasiprobabilities, labeled by $a = a'$, equal each other:

$$\text{Tr}\big(\Pi_{\max}\,\rho(\theta)\,\Pi_{\max}\,K_+^\dagger K_+\big) = \text{Tr}\big(\Pi_{\min}\,\rho(\theta)\,\Pi_{\min}\,K_+^\dagger K_+\big). \tag{D4}$$

This condition is satisfied by the main text's qubit system of interest, $A = \sigma_x/2$, and input state $(|a_+\rangle + |a_-\rangle)/2 = |0\rangle$.

Generalized equality: The postselected QFI, $\mathcal{I}(\theta)$, is proportional to the quasiprobability's nonclassicality gap:

$$\mathcal{I}(\theta) = 4\,(a_{\max} - a_{\min})^2 \left(\max_{a,a'}\big\{|\tilde{p}_{\rho(\theta)}(a, a'|+)|^2\big\} - \min_{a,a'}\big\{|\tilde{p}_{\rho(\theta)}(a, a'|+)|^2\big\} \right). \tag{D5}$$

Proof: We calculate postselected quantum Fisher information by substituting from Eq. (D1) into a formula for a pure state's postselected QFI {Eq. (5) in [15]}:

$$\mathcal{I}(\theta) = \frac{4}{p^{\text{ps}}(\theta)}\,\text{Tr}\big(A\,\rho(\theta)\,A\,K_+^\dagger K_+\big) - \frac{4}{[p^{\text{ps}}(\theta)]^2}\,|\text{Tr}\big(A\,\rho(\theta)\,K_+^\dagger K_+\big)|^2 \tag{D6}$$

$$= \frac{4}{[p^{\text{ps}}(\theta)]^2}\,\Big[\text{Tr}\big(A\,\rho(\theta)\,A\,K_+^\dagger K_+\big)\,\text{Tr}\big(\rho(\theta)\,K_+^\dagger K_+\big) - \text{Tr}\big(A\,\rho(\theta)\,K_+^\dagger K_+\big)\,\text{Tr}\big(\rho(\theta)\,A\,K_+^\dagger K_+\big)\Big]. \tag{D7}$$

We will rewrite the right-hand side, by decomposing the evolved state as

$$\rho(\theta) = (\Pi_{\max} + \Pi_{\min})\,\rho(\theta)\,(\Pi_{\max} + \Pi_{\min}) \tag{D8}$$

$$= \Pi_{\max}\,\rho(\theta)\,\Pi_{\max} + \Pi_{\max}\,\rho(\theta)\,\Pi_{\min} + \Pi_{\min}\,\rho(\theta)\,\Pi_{\max} + \Pi_{\min}\,\rho(\theta)\,\Pi_{\min}. \tag{D9}$$

We substitute this expression into Eq. (D7). Then, we invoke the eigenvalue equations $A\Pi_{\max} = a_{\max}\Pi_{\max}$ and $A\Pi_{\min} = a_{\min}\Pi_{\min}$. Once we multiply out, terms cancel:

$$\mathcal{I}(\theta|\rho(\theta)) = \frac{4}{[p^{\text{ps}}(\theta)]^2}\,(a_{\max} - a_{\min})^2\big[\text{Tr}\big(\Pi_{\max}\,\rho(\theta)\,\Pi_{\max}\,K_+^\dagger K_+\big)\,\text{Tr}\big(\Pi_{\min}\,\rho(\theta)\,\Pi_{\min}\,K_+^\dagger K_+\big)$$
$$- \text{Tr}\big(\Pi_{\max}\,\rho(\theta)\,\Pi_{\min}\,K_+^\dagger K_+\big)\,\text{Tr}\big(\Pi_{\min}\,\rho(\theta)\,\Pi_{\max}\,K_+^\dagger K_+\big)\big]. \tag{D10}$$

The final term equals $|\text{Tr}\big(\Pi_{\max}\,\rho(\theta)\,\Pi_{\min}\,K_+^\dagger K_+\big)|^2$, so

$$\mathcal{I}(\theta) = 4\,(a_{\max} - a_{\min})^2\,\big[\tilde{p}_{\rho(\theta)}(a_{\max}, a_{\max}|+)\,\tilde{p}_{\rho(\theta)}(a_{\min}, a_{\min}|+) - |\tilde{p}_{\rho(\theta)}(a_{\max}, a_{\min}|+)|^2\big]. \tag{D11}$$

Since $\mathcal{I}(\theta)$ and $(a_{\max} - a_{\min})^2 \geq 0$, Eq. (D11) implies

$$\tilde{p}_{\rho(\theta)}(a_{\max}, a_{\max}|+)\,\tilde{p}_{\rho(\theta)}(a_{\min}, a_{\min}|+) \geq |\tilde{p}_{\rho(\theta)}(a_{\max}, a_{\min}|+)|^2. \tag{D12}$$

The left-hand side equals $\big[\tilde{p}_{\rho(\theta)}(a_{\max}, a_{\max}|+)\big]^2$, by Eq. (D4). Since each side of Eq. (D4) is overtly a probability, $\tilde{p}_{\rho(\theta)}(a_{\max}, a_{\max}|+)$ is real. The square therefore equals the square modulus. Substituting into Ineq. (D12) yields $|\tilde{p}_{\rho(\theta)}(a_{\max}, a_{\max}|+)|^2 \geq |\tilde{p}_{\rho(\theta)}(a_{\max}, a_{\min}|+)|^2$. This inequality contains the square moduli of all the quasiprobabilities for which $f = +$. Hence the inequality's left-hand side equals the greatest square probability, $\max_{a,a'}\big\{[\tilde{p}_{\rho(\theta)}(a, a'|+)]^2\big\}$, while the right-hand side equals the least square modulus, $\min_{a,a'}\big\{[\tilde{p}_{\rho(\theta)}(a, a'|+)]^2\big\}$. Substituting the max and min into Eq. (D11) yields Eq. (D5). $\square$

REFERENCES

[1] L. A. Rozema, A. Darabi, D. H. Mahler, A. Hayat, Y. Soudagar, and A. M. Steinberg, Violation of heisenberg's measurement-disturbance relationship by weak measurements, Phys. Rev. Lett. **109**, 100404 (2012).

[2] V. Giovannetti, S. Lloyd, and L. Maccone, Quantum-enhanced measurements: beating the standard quantum limit, Science **306**, 1330 (2004).

[3] V. Giovannetti, S. Lloyd, and L. Maccone, Quantum metrology, Physical review letters **96**, 010401 (2006).

[4] V. Giovannetti, S. Lloyd, and L. Maccone, Advances in quantum metrology, Nature photonics **5**, 222 (2011).

[5] E. Polino, M. Valeri, N. Spagnolo, and F. Sciarrino, Photonic quantum metrology, AVS Quantum Science **2**, 024703 (2020), https://doi.org/10.1116/5.0007577.

[6] S.-J. Yoon, J.-S. Lee, C. Rockstuhl, C. Lee, and K.-G. Lee, Experimental quantum polarimetry using heralded single photons, Metrologia **57**, 045008 (2020).

[7] J. Abadie *et al.* (LIGO Scientific Collaboration), A gravitational wave observatory operating beyond the quantum shot-noise limit, Nature Physics 2011 7:12 **7**, 962 (2011).

[8] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Observation of gravitational waves from a binary black hole merger, Phys. Rev. Lett. **116**, 061102 (2016).

[9] N. Ghosh and A. I. Vitkin, Tissue polarimetry: concepts, challenges, applications, and outlook, Journal of biomedical optics **16**, 110801 (2011).

[10] D. Budker and M. Romalis, Optical magnetometry, Nature physics **3**, 227 (2007).

[11] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, Randomized benchmarking of quantum gates, Physical Review A **77**, 012307 (2008).

[12] M. Dobšíček, G. Johansson, V. Shumeiko, and G. Wendin, Arbitrary accuracy iterative quantum phase estimation algorithm using a single ancillary qubit: A two-qubit benchmark, Physical Review A **76**, 030306 (2007).

[13] H. Cramér, *Mathematical methods of statistics (PMS-9)*, Vol. 9 (Princeton University Press, 2016).

[14] C. R. Rao, Information and the accuracy attainable in the estimation of statistical parameters, in *Breakthroughs in statistics* (Springer, 1992) pp. 235–247.

[15] D. R. M. Arvidsson-Shukur, N. Yunger Halpern, H. V. Lepage, A. A. Lasek, C. H. W. Barnes, and S. Lloyd, Quantum advantage in postselected metrology, Nature Communications **11**, 3775 (2020).

[16] J. Combes, C. Ferrie, Z. Jiang, and C. M. Caves, Quantum limits on postselected, probabilistic quantum metrology, Phys. Rev. A **89**, 052117 (2014).

[17] C. Ferrie and J. Combes, Weak value amplification is suboptimal for estimation and detection, Phys. Rev. Lett. **112**, 040406 (2014).

[18] J. Harris, R. W. Boyd, and J. S. Lundeen, Weak value amplification can outperform conventional measurement in the presence of detector saturation, Phys. Rev. Lett. **118**, 070802 (2017).

[19] J. Sinclair, M. Hallaji, A. M. Steinberg, J. Tollaksen, and A. N. Jordan, Weak-value amplification and optimal parameter estimation in the presence of correlated noise, Phys. Rev. A **96**, 052128 (2017).

[20] M. Hallaji, A. Feizpour, G. Dmochowski, J. Sinclair, and A. M. Steinberg, Weak-value amplification of the nonlinear effect of a single photon, Nature Physics **13**, 540 (2017).

[21] G. I. Viza, J. Martínez-Rincón, G. B. Alves, A. N. Jordan, and J. C. Howell, Experimentally quantifying the advantages of weak-value-based metrology, Phys. Rev. A **92**, 032127 (2015).

[22] X. Qiu, L. Xie, X. Liu, L. Luo, Z. Li, Z. Zhang, and J. Du, Precision phase estimation based on weak-value amplification, Applied Physics Letters **110**, 071105 (2017).

[23] Y. Aharonov, D. Z. Albert, and L. Vaidman, How the result of a measurement of a component of the spin of a spin- *1/2* particle can turn out to be 100, Phys. Rev. Lett. **60**, 1351 (1988).

[24] I. M. Duck, P. M. Stevenson, and E. C. G. Sudarshan, The sense in which a "weak measurement" of a spin-1/2 particle's spin component yields a value 100, Phys. Rev. D **40**, 2112 (1989).

[25] O. Hosten and P. Kwiat, Observation of the spin hall effect of light via weak measurements, Science **319**, 787 (2008).

[26] P. B. Dixon, D. J. Starling, A. N. Jordan, and J. C. Howell, Ultrasensitive beam deflection measurement via interferometric weak value amplification, Phys. Rev. Lett. **102**, 173601 (2009).

[27] S. Pang, J. Dressel, and T. A. Brun, Entanglement-assisted weak value amplification, Phys. Rev. Lett. **113**, 030401 (2014).

[28] S. Pang and T. A. Brun, Improving the precision of weak measurements by postselection measurement, Phys. Rev. Lett. **115**, 120401 (2015).

[29] A. N. Jordan, J. Martínez-Rincón, and J. C. Howell, Technical advantages for weak-value amplification: when less is more, Physical Review X **4**, 011031 (2014).

[30] D. J. Starling, P. B. Dixon, A. N. Jordan, and J. C. Howell, Optimizing the signal-to-noise ratio of a beam-deflection measurement with interferometric weak values, Phys. Rev. A **80**, 041803 (2009).

[31] D. J. Starling, P. B. Dixon, A. N. Jordan, and J. C. Howell, Precision frequency measurements with interferometric weak values, Phys. Rev. A **82**, 063822 (2010).

[32] O. S. Magaña Loaiza, M. Mirhosseini, B. Rodenburg, and R. W. Boyd, Amplification of angular rotations using weak measurements, Phys. Rev. Lett. **112**, 200401 (2014).

[33] K. Lyons, J. Dressel, A. N. Jordan, J. C. Howell, and P. G. Kwiat, Power-recycled weak-value-based metrology, Phys. Rev. Lett. **114**, 170801 (2015).

[34] J. Martínez-Rincón, C. A. Mullarkey, G. I. Viza, W.-T. Liu, and J. C. Howell, Ultrasensitive inverse weak-value tilt meter, Opt. Lett. **42**, 2479 (2017).

[35] P. Egan and J. A. Stone, Weak-value thermostat with 0.2 mk precision, Opt. Lett **37**, 4991 (2012).

[36] H. F. Hofmann, M. E. Goggin, M. P. Almeida, and M. Barbieri, Estimation of a quantum interaction parameter using weak measurements: Theory and experiment, Phys. Rev. A **86**, 040102 (2012).

[37] J. G. Kirkwood, Quantum statistics of almost classical assemblies, Phys. Rev. **44**, 31 (1933).

[38] P. A. M. Dirac, On the analogy between classical and quantum mechanics, Rev. Mod. Phys. **17**, 195 (1945).

[39] L. M. Johansen, Quantum theory of successive projective measurements, Phys. Rev. A **76**, 012119 (2007).

[40] J. S. Lundeen, B. Sutherland, A. Patel, C. Stewart, and C. Bamber, Direct measurement of the quantum wavefunction, Nature **474**, 188 (2011).

[41] J. S. Lundeen and C. Bamber, Procedure for direct measurement of general quantum states using weak measurement, Phys. Rev. Lett. **108**, 070402 (2012).

[42] C. Bamber and J. S. Lundeen, Observing dirac's classical phase space analog to the quantum state, Phys. Rev. Lett. **112**, 070405 (2014).

[43] G. S. Thekkadath, L. Giner, Y. Chalich, M. J. Horton, J. Banker, and J. S. Lundeen, Direct measurement of the density matrix of a quantum system, Phys. Rev. Lett. **117**, 120401 (2016).

[44] N. Yunger Halpern, Jarzynski-like equality for the out-of-time-ordered correlator, Phys. Rev. A **95**, 012120 (2017).

[45] N. Yunger Halpern, B. Swingle, and J. Dressel, Quasiprobability behind the out-of-time-ordered correlator, Phys. Rev. A **97**, 042105 (2018).

[46] J. González Alonso, N. Yunger Halpern, and J. Dressel, Out-of-time-ordered-correlator quasiprobabilities robustly witness scrambling, Phys. Rev. Lett. **122**, 040404 (2019).

[47] N. Y. Halpern, A. Bartolotta, and J. Pollack, Entropic uncertainty relations for quantum information scrambling, Communications Physics **2**, 1 (2019).

[48] R. Mohseninia, J. R. G. Alonso, and J. Dressel, Optimizing measurement strengths for qubit quasiprobabilities behind out-of-time-ordered correlators, Phys. Rev. A **100**, 062336 (2019).

[49] A. M. Steinberg, Conditional probabilities in quantum theory and the tunneling-time controversy, Phys. Rev. A **52**, 32 (1995).

[50] J. Dressel, M. Malik, F. M. Miatto, A. N. Jordan, and R. W. Boyd, Colloquium: Understanding quantum weak values: Basics and applications, Rev. Mod. Phys. **86**, 307 (2014).

[51] M. F. Pusey, Anomalous weak values are proofs of contextuality, Phys. Rev. Lett. **113**, 200401 (2014).

[52] D. R. M. Arvidsson-Shukur, A. N. O. Gottfries, and C. H. W. Barnes, Evaluation of counterfactuality in counterfactual communication protocols, Phys. Rev. A **96**, 062316 (2017).

[53] D. R. M. Arvidsson-Shukur and C. H. W. Barnes, Postselection and counterfactual communication, Phys. Rev. A **99**, 060102 (2019).

[54] R. Kunjwal, M. Lostaglio, and M. F. Pusey, Anomalous weak values and contextuality: Robustness, tightness, and imaginary parts, Phys. Rev. A **100**, 042116 (2019).

[55] J. H. Jenne and D. R. M. Arvidsson-Shukur, Quantum learnability is arbitrarily distillable, arXiv preprint arXiv:1909.11116 (2021).

[56] R. Jozsa, Complex weak values in quantum measurement, Phys. Rev. A **76**, 044103 (2007).

[57] H. F. Hofmann, On the role of complex phases in the quantum statistics of weak measurements, New Journal of Physics **13**, 103009 (2011).

[58] J. Dressel and A. N. Jordan, Significance of the imaginary part of the weak value, Phys. Rev. A **85**, 012107 (2012).

[59] J. T. Monroe, N. Yunger Halpern, T. Lee, and K. W. Murch, Weak measurement of a superconducting qubit reconciles incompatible operators, Phys. Rev. Lett. **126**, 100403 (2021).

[60] A. E. Allahverdyan, Nonequilibrium quantum fluctuations of work, Phys. Rev. E **90**, 032137 (2014).

[61] H. J. D. Miller and J. Anders, Time-reversal symmetric work distributions for closed quantum dynamics in the histories framework, New Journal of Physics **19**, 062001 (2017).

[62] A. Levy and M. Lostaglio, A quasiprobability distribution for heat fluctuations in the quantum regime, arXiv preprint arXiv:1909.11116 (2019).

[63] M. Lostaglio, Certifying quantum signatures in thermodynamics and metrology via contextuality of quantum linear response, Phys. Rev. Lett. **125**, 230603 (2020).

[64] R. B. Griffiths, Consistent histories and the interpretation of quantum mechanics, Journal of Statistical Physics **36**, 219 (1984).

[65] S. Goldstein and D. N. Page, Linearly positive histories: Probabilities for a robust family of sequences of quantum events, Phys. Rev. Lett. **74**, 3715 (1995).

[66] J. B. Hartle, Linear positivity and virtual probability, Phys. Rev. A **70**, 022104 (2004).

[67] H. F. Hofmann, Complex joint probabilities as expressions of reversible transformations in quantum mechanics, New Journal of Physics **14**, 043031 (2012).

[68] H. F. Hofmann, Derivation of quantum mechanics from a single fundamental modification of the relations between physical properties, Phys. Rev. A **89**, 042115 (2014).

[69] H. F. Hofmann, Quantum paradoxes originating from the nonclassical statistics of physical properties related to each other by half-periodic transformations, Phys. Rev. A **91**, 062123 (2015).

[70] H. F. Hofmann, On the fundamental role of dynamics in quantum physics, The European Physical Journal D **70**, 118 (2016).

[71] J. J. Halliwell, Leggett-garg inequalities and no-signaling in time: A quasiprobability approach, Phys. Rev. A **93**, 022123 (2016).

[72] B. C. Stacey, Quantum theory as symmetry broken by vitality, arXiv preprint arXiv:1907.02432 (2019).

[73] D. R. M. Arvidsson-Shukur, J. Chevalier Drori, and N. Yunger Halpern, Conditions tighter than noncommutation needed for nonclassicality, Journal of Physics A: Mathematical and Theoretical **54**, 284001 (2021).

[74] S. De Bievre, Kirkwood-Dirac nonclassicality, support uncertainty and complete incompatibility, arXiv e-prints , arXiv:2106.10017 (2021), arXiv:2106.10017 [quant-ph].

[75] M. Oszmaniec, D. J. Brod, and E. F. Galvão, Measuring relational information between quantum states, and applications, arXiv e-prints , arXiv:2109.10006 (2021), arXiv:2109.10006 [quant-ph].

[76] S. L. Braunstein and C. M. Caves, Statistical distance and the geometry of quantum states, Phys. Rev. Lett. **72**, 3439

(1994).

[77] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*, 10th ed. (Cambridge University Press, New York, NY, USA, 2011).

[78] H. M. Wiseman, Weak values, quantum trajectories, and the cavity-qed experiment on wave-particle correlation, Phys. Rev. A **65**, 032111 (2002).