

Certainty in Heisenberg's uncertainty principle: Revisiting definitions for estimation errors and disturbance

Justin Dressel^{1,*} and Franco Nori^{1,2}¹Center for Emergent Matter Science, RIKEN, Saitama 351-0198, Japan²Physics Department, University of Michigan, Ann Arbor, Michigan 48109-1040, USA

(Received 22 August 2013; published 10 February 2014)

We revisit the definitions of error and disturbance recently used in error-disturbance inequalities derived by Ozawa and others by expressing them in the reduced system space. The interpretation of the definitions as mean-squared deviations relies on an implicit assumption that is generally incompatible with the Bell-Kochen-Specker-Spekkens contextuality theorems, and which results in averaging the deviations over a non-positive-semidefinite joint quasiprobability distribution. For unbiased measurements, the error admits a concrete interpretation as the dispersion in the estimation of the mean induced by the measurement ambiguity. We demonstrate how to directly measure not only this dispersion but also every observable moment with the same experimental data, and thus demonstrate that perfect distributional estimations can have nonzero error according to this measure. We conclude that the inequalities using these definitions do not capture the spirit of Heisenberg's eponymous inequality, but do indicate a qualitatively different relationship between dispersion and disturbance that is appropriate for ensembles being probed by all outcomes of an apparatus. To reconnect with the discussion of Heisenberg, we suggest alternative definitions of error and disturbance that are intrinsic to a single apparatus outcome. These definitions naturally involve the retrodictive and interdictive states for that outcome, and produce complementarity and error-disturbance inequalities that have the same form as the traditional Heisenberg relation.

DOI: 10.1103/PhysRevA.89.022106

PACS number(s): 03.65.Ta, 03.67.-a, 02.50.Cw, 03.65.Fd

I. INTRODUCTION

One of the fundamental principles of quantum mechanics is the uncertainty principle, which places a restriction upon the degree to which one can constrain the likelihoods of future measurements made on a quantum system. The initial form of this principle was postulated by Heisenberg [1] and rigorously derived by Kennard [2]. Heisenberg subsequently corrected his original idea and presented its refined form in a remarkable series of lectures [3]. The modern form of this principle extends the work of Heisenberg by directly relating this restriction to the noncommutativity of the algebra of quantum observable operators. The abstract algebraic generalization was first derived for pure quantum states by Weyl and Robertson [4,5], but equally applies to other mathematical contexts (e.g., Fourier analysis).

A precise statement of the uncertainty principle from the quantum perspective is the following: If an experimenter repeatedly prepares a system in a particular quantum state $\hat{\rho}$ and subsequently measures two observables \hat{A} and \hat{B} of the system using *independent* preparations, then the accumulated statistics of the measurement will display variances, $\sigma_A^2 = \langle (\hat{A} - \langle \hat{A} \rangle)^2 \rangle$ and $\sigma_B^2 = \langle (\hat{B} - \langle \hat{B} \rangle)^2 \rangle$, that must satisfy the inequality

$$\sigma_A \sigma_B \geq \left| \frac{1}{2i} \langle [\hat{A}, \hat{B}] \rangle \right|, \quad (1)$$

where $\langle \hat{A} \rangle = \text{Tr}[\hat{A} \hat{\rho}]$ is the usual expectation value, and $[\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A}$ is the commutator of the observable operators. This inequality follows in turn from the stronger Cauchy-Schwartz inequality that must hold for any operators

 \hat{A} and \hat{B} ,

$$\sigma_A^2 \sigma_B^2 \geq [\langle \hat{A} * \hat{B} \rangle - \langle \hat{A} \rangle \langle \hat{B} \rangle]^2 + \left| \frac{1}{2i} \langle [\hat{A}, \hat{B}] \rangle \right|^2, \quad (2)$$

and which was derived for a pure quantum state $\hat{\rho} = |\psi\rangle\langle\psi|$ by Schrödinger [6]. Here we write it in terms of the symmetric Jordan product [7] of the operators $\hat{A} * \hat{B} = (\hat{A}\hat{B} + \hat{B}\hat{A})/2$, which will appear later in our discussion.

Recently it has become clear that the inequality in Eq. (1) is commonly (and perhaps incorrectly) associated with three conceptually distinct statements:

(1) A preparation $\hat{\rho}$ has intrinsic spreads in \hat{A} and \hat{B} such that Eq. (1) is satisfied for independent measurements.

(2) Estimating both \hat{A} and \hat{B} simultaneously will show measured estimation errors that satisfy Eq. (1).

(3) Estimating \hat{A} will disturb subsequent estimations of \hat{B} such that the measured estimation error of \hat{A} and disturbance of \hat{B} satisfy Eq. (1).

All three of these statements relate to the original discussion by Heisenberg in Ref. [1], but only the first statement corresponds to the derivation of Eq. (1).

The remaining two statements correspond to experimental situations that do not relate to Eq. (1) directly, and that contain undefined new concepts. The second statement concerns the complementarity (in the sense of Bohr) between the simultaneous *estimation errors* for observables \hat{A} and \hat{B} . The third statement concerns the trade-off between the estimation error of \hat{A} and the resulting *disturbance* detectable by a subsequent measurement of \hat{B} .

Finding appropriate definitions (and resulting inequalities) that describe these distinct concepts of estimation error and disturbance has had a long history, much of which has been carefully reviewed by Busch *et al.* [8]. This pursuit can be traced back to the work of Arthurs and Kelly [9], who first demonstrated that a naive application of Eq. (1)

*Department of Electrical Engineering, University of California, Riverside, California 92521, USA.

will not describe the complementarity between estimations of a particular preparation state performed by a particular measuring apparatus. They used a definition of the estimation error that generalizes the notion of classical mean-squared error for individual measurements [10,11] (and which we will analyze more carefully later in this paper). Many papers subsequently appeared that highlighted different special cases where inequalities formally mimicking Eq. (1) were inadequate for certain classes of states or observables using similar definitions that built off the work of Arthurs and Kelly (e.g., [12–20]).

A paradigm shift occurred when Ozawa [21] derived a *universal* inequality that was valid for any preparation state and a generic coupling to a detector

$$\epsilon_A \eta_B + \epsilon_A \sigma_B + \sigma_A \eta_B \geq \left| \frac{1}{2i} \langle [\hat{A}, \hat{B}] \rangle \right| \quad (3)$$

using formal definitions for the mean-squared error ϵ_A and disturbance η_B also suggested by Appleby [17,18] and based directly on the work of Arthurs and Kelly [9]. Shortly thereafter, Hall and Ozawa independently derived a similar universal inequality for the complementarity of joint observable estimations [22,23]

$$\epsilon_A \epsilon_B + \epsilon_A \sigma_B + \sigma_A \epsilon_B \geq \left| \frac{1}{2i} \langle [\hat{A}, \hat{B}] \rangle \right|, \quad (4)$$

building off Ozawa's previous result [24].

Due to the similarity between these inequalities and Eq. (1), as well as their resemblance to the discussion in Ref. [1], Eqs. (3) and (4) have since been marketed as *corrections* to Heisenberg's eponymous uncertainty principle that properly address the distinct concepts of estimation error and disturbance. Nevertheless, these inequalities are still derived from Eq. (1) in its role as a generic operator inequality by making judicious replacements of \hat{A} and \hat{B} . As such, the interpretation of these new inequalities as corrections to the uncertainty principle crucially depends upon the physical significance of the additional quantities ϵ_A and η_B , and whether their definitions adequately reflect the situation considered by Heisenberg.

The definitions of the estimation error ϵ_A and the disturbance η_B used by the inequalities in Eqs. (3) and (4) can be understood as quantizations of classical mean-squared error and disturbance definitions [11]. Classically, these definitions involve the second moment of the difference between each measured result and well-defined reference values; the reference values are considered to be "correct" values, so that the differences may be interpreted as the "errors" of each individual measurement. In contrast, the quantized versions formally involve the second moment of a difference between joint system-detector operators in the Heisenberg picture. There are generally no well-defined reference values for each individual quantum measurement, so these formal definitions are the closest quantized equivalents to the classical definitions.

Due to the formal nature of these definitions, and in spite of their correspondence to classical definitions, there has been controversy regarding both the physical significance of ϵ_A and η_B in a quantum setting, as well as the feasibility of experimentally measuring them. Indeed, there have been several independent appeals to find alternative and operationally motivated definitions [8,25–33] that produce inequalities which faithfully reflect Heisenberg's original discussion as presented in Ref. [1], and that also have a form similar to the inequality in Eq. (1).

To address this controversy, Ozawa proposed an indirect method to experimentally determine the mean-squared error and disturbance [34], which has recently been implemented by Erhart *et al.* [35], Sulyok *et al.* [36], and Baek *et al.* [37] using neutron-optical setups. This method involves the preparation and measurement of three distinct but related states in order to formally construct the mean-squared error and disturbance associated with the measurement of one of those states.

To improve upon this indirect method by removing the need for three related preparations, Lund and Wiseman [38] independently proposed an alternative procedure for measuring the mean-squared error and disturbance that requires only a single preparation state. By expressing the error and disturbance in terms of a joint Terletsky-Margenau-Hill [39,40] quasiprobability distribution, they can be related [41] to weak values [42,43] that may be approximately measured by a separate weakly correlated detector [44]. This alternative procedure was recently implemented by Rozema *et al.* [45] and Kaneda *et al.* [46] using optical setups.

A similar quasiprobability technique was recently proposed by Weston *et al.* [47] to derive (and then experimentally test) a tighter universal inequality for complementarity

$$\epsilon_A \frac{\sigma_B + \sigma_{B,\text{est}}}{2} + \epsilon_B \frac{\sigma_A + \sigma_{A,\text{est}}}{2} \geq \left| \frac{1}{2i} \langle [\hat{A}, \hat{B}] \rangle \right|, \quad (5)$$

where $\sigma_{j,\text{est}}$ represents the measured standard deviation of the estimation. This tighter inequality has since been supplemented by inequalities derived by Branciard [48,49],

$$\sigma_A^2 \epsilon_B^2 + \epsilon_A^2 \sigma_B^2 + 2 \epsilon_A^2 \epsilon_B^2 \sqrt{\sigma_A^2 \sigma_B^2 - C_{AB}^2} \geq C_{AB}^2, \quad (6)$$

where $C_{AB} = |\langle [\hat{A}, \hat{B}] \rangle / 2i|$. Substituting η_B for ϵ_B in Eq. (6) produces the tighter error-disturbance inequality corresponding to Eq. (3). The Branciard inequalities have also been recently tested experimentally by Kaneda *et al.* [46].

In this paper we take a step back from this rapid progression to reassess the formal definitions of the estimation error ϵ and the measurement disturbance η used by the inequalities in Eqs. (3)–(6). To be clear, we do not challenge these results, since they are assuredly important and thought provoking in their own right. Instead, we explore the question of whether these results faithfully correct the idea of the uncertainty principle as discussed by Heisenberg in Refs. [1,3], or whether they provide a qualitatively different and supplementary understanding to that principle. By examining the definitions used in these inequalities from an instrumental approach [44,50–53], we illustrate several subtle and unsatisfactory features that seem to deviate from the intent of Heisenberg.

Despite the obvious correspondence to the classical ideas of mean-squared error and disturbance, the interpretation of the formal definitions of ϵ^2 and η^2 as "mean-squared deviations" relies on an implicit assumption that observables can be assigned definite values before they are measured. This assumption is in direct violation of the Bell-Kochen-Specker-Spekkens contextuality theorems [54–58], and mandates the use of non-positive-definite quasiprobability distributions for interpreting the definitions as average deviations [59]. Unlike classical observables, quantum observables generally do not have well-defined values prior to measurement from which one can construct meaningful deviations for each measurement

realization. Heisenberg was careful to avoid this appeal to any definite reference values [3].

In the special case of an unbiased measurement (i.e., a measurement that faithfully determines the mean for any initial state), the quantity ϵ can be given a realistic statistical interpretation, though not as a root-mean-squared deviation. Instead, it is the added *dispersion* of the mean of an ensemble of measurements that stems from the ambiguity of the estimation. However, we emphasize that ϵ does not quantify the error in the estimation of the full distribution for the observable, but only the dispersion in the estimation of its mean. Indeed, we show how all moments of the estimated observable (and thus its full distribution) may be determined in a single experimental run even when ϵ is nonzero. If the detector happens to report the same value range as the measured observable, then one can loosely interpret such a dispersion as an average deviation of each estimation from any possible reference value of the observable, but this loose interpretation fails for more general detector outputs that do not directly correspond to the spectral range of the measured observable. We also highlight the relation between the mean-squared disturbance η and the average Lindblad perturbation to the state that is induced by the estimating apparatus, and show that η does not quantify this measurable perturbation in a natural way.

To reconnect with the original discussion of Heisenberg, we suggest an alternative perspective on estimation error and disturbance that focuses on what can be inferred on average from a *single* apparatus outcome, rather than what can be inferred on average from *all* apparatus outcomes. Using alternative definitions of the estimation error and disturbance founded on recent work on the retrodictive and interdictive states associated with a single apparatus outcome [53], we derive and generalize inequalities that were independently obtained by Hofmann [25]. These inequalities also have the form of Eq. (1), but correspond to the complementarity and error-disturbance interpretations of the Heisenberg uncertainty relation as it relates to a single apparatus outcome. This result complements similar results obtained by Busch *et al.* [32,33] and Watanabe *et al.* [29,30], who also use different definitions for error and disturbance that are intrinsic to the apparatus.

This paper is organized as follows. In Sec. II we briefly review the instrumental formalism of indirect observable measurement to keep the discussion self-contained. In Sec. III we examine the definition of estimation error ϵ as used by the inequalities of Eqs. (3)–(6). In Sec. IV we examine the corresponding definition of measurement disturbance η . In Sec. V we consider alternative definitions of estimation error and disturbance that depend only on a single apparatus outcome, and consequently produce an alternative set of inequalities that have the same form as the traditional Heisenberg relation. We conclude in Sec. VI.

II. INDIRECT MEASUREMENT

The universal inequalities in Eqs. (3)–(6) all pertain to indirect observable measurements made using all outcomes of a detecting apparatus. We will find it useful in what follows to discuss these sorts of measurements using a *quantum instrument* approach [50,51] augmented by the *contextual-values* formalism [44,52,53]. For completeness, we now

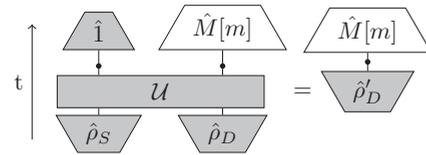


FIG. 1. Detector picture of an indirect measurement, highlighted in gray, with time increasing along the vertical axis. (See [53] for a more detailed description of the graphical notation used throughout this paper.) An initially uncorrelated system state $\hat{\rho}_S$ and detector state $\hat{\rho}_D$ are coupled with a unitary interaction \mathcal{U} . After the reference time indicated by the black dots, a joint observable $\hat{I} \otimes \hat{M}[m]$ with spectral function m_k over the detector outcomes indexed by k is projectively measured. No further measurement is considered on the system space, which is indicated by the identity \hat{I} in the joint observable $\hat{I} \otimes \hat{M}[m]$. This procedure is formally equivalent to the measurement of $\hat{M}[m]$ directly after the preparation of a reduced detector state $\hat{\rho}'_D$ that contains correlated information about the initial system state $\hat{\rho}_S$.

briefly review this approach in both the detector picture (where the system is traced out) and the system picture (where the detector is traced out).

A. Detector picture

A system state—represented most generally by a density operator $\hat{\rho}_S$ —is coupled to an uncorrelated [60] detector state $\hat{\rho}_D$ with an interaction \mathcal{U} characterized by a unitary operator \hat{U} , as illustrated schematically in Fig. 1. The joint state after the coupling is then correlated $\hat{\rho}'_{SD} = \mathcal{U}(\hat{\rho}_S \otimes \hat{\rho}_D) = \hat{U}(\hat{\rho}_S \otimes \hat{\rho}_D)\hat{U}^\dagger$. Performing a partial trace over the system produces the reduced detector state $\hat{\rho}'_D = \text{Tr}_S[\hat{\rho}'_{SD}]$, which now contains information about the initial system state. To exploit this correlation, the detector is read (i.e., measured projectively in some basis $|k\rangle$) and an eigenvalue m_k is assigned by the experimenter to each detector outcome k . After many identical trials, the average of the recorded values converges to the sum $\sum_k m_k p_k$, where $p_k = \langle k | \hat{\rho}'_D | k \rangle$ are the probabilities for observing the measured detector outcomes.

This empirical procedure corresponds to constructing a particular detector observable $\hat{M}[m] = \sum_k m_k |k\rangle\langle k|$, and measuring it projectively with respect to the reduced postinteraction detector state $\text{Tr}_D[\hat{M}[m]\hat{\rho}'_D]$. The chosen set of values $\{m_k\}$ assigned to each detector outcome k forms the spectral function m for the observable $\hat{M}[m]$. We keep the dependence of this observable on the chosen spectral function m explicit, since its role in what follows will be important.

For brevity hereafter, we will use standard condensed notation for the joint expectation value $\langle \cdot \rangle = \text{Tr}[(\cdot)(\hat{\rho}_S \otimes \hat{\rho}_D)]$, the detector expectation value $\langle \cdot \rangle_D = \text{Tr}_D[(\cdot)\hat{\rho}_D]$, the postinteraction detector expectation value $\langle \cdot \rangle_{D'} = \text{Tr}_D[(\cdot)\hat{\rho}'_D]$, the system expectation value $\langle \cdot \rangle_S = \text{Tr}_S[(\cdot)\hat{\rho}_S]$, and the postinteraction system expectation value $\langle \cdot \rangle_{S'} = \text{Tr}_S[(\cdot)\hat{\rho}'_S]$.

B. System picture

One can also perform a partial trace over the detector to produce a different picture of this empirical procedure that is contained entirely in the system space, as shown in Fig. 2, and which will be more illuminating for what follows. In

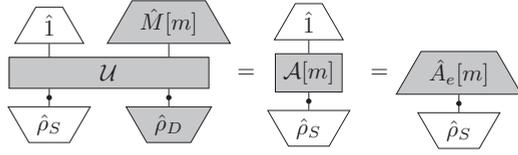


FIG. 2. System picture of the same indirect measurement as in Fig. 1, highlighted in gray. The joint measurement procedure is equivalent to a *quantum instrument* $\mathcal{A}[m]$ (see Sec. II B) that acts directly on the system after a reference time indicated by the black dots. Since no subsequent measurements are considered (as indicated by the identity operator $\hat{1}$), then this instrument induces a weighted POM $\hat{A}_e[m]$ as the system observable that is effectively measured after the preparation of the initial state $\hat{\rho}_S$, as shown in Eq. (9).

this picture, observing a particular outcome k on the detector induces a *quantum operation* \mathcal{A}_k on the system

$$\begin{aligned} \mathcal{A}_k(\hat{\rho}_S) &= \langle k | \hat{U}(\hat{\rho}_S \otimes \hat{\rho}_D) \hat{U}^\dagger | k \rangle \\ &= \sum_l p_l \langle k | \hat{U} | l \rangle \hat{\rho}_S \langle l | \hat{U}^\dagger | k \rangle \\ &= \sum_l \hat{M}_{k,l} \hat{\rho}_S \hat{M}_{k,l}^\dagger, \end{aligned} \quad (7)$$

where each $\hat{M}_{k,l} = \sqrt{p_l} \langle k | \hat{U} | l \rangle$ is a Kraus operator that characterizes the operation, and $\hat{\rho}_D = \sum_l p_l |l\rangle \langle l|$ is some (nonunique) pure state decomposition of the initial detector state. It follows that the procedure for measuring the detector observable $\hat{M}[m]$ equivalently weights a set of operations being performed on the system to produce a *quantum instrument*

$$\mathcal{A}[m] = \sum_k m_k \mathcal{A}_k, \quad (8)$$

that completely describes the action of the indirect detector [50,51,53,61]. Intuitively, this instrument is the mathematical representation of the laboratory apparatus that is making the measurement.

Averaging the recorded values m_k chosen by the experimenter therefore produces the set of formal equalities

$$\begin{aligned} \sum_k m_k p_k &= \langle \hat{U}^\dagger(\hat{1} \otimes \hat{M}[m]) \hat{U} \rangle = \langle \hat{M}[m] \rangle_D \\ &= \text{Tr}_S[\mathcal{A}[m](\hat{\rho}_S)] = \langle \hat{A}_e[m] \rangle_S \end{aligned} \quad (9)$$

that are illustrated in Figs. 1 and 2. The first equality is the joint system-detector picture that includes the interaction \hat{U} and the joint observable $\hat{1} \otimes \hat{M}[m]$. The second equality is the detector picture involving the detector observable $\hat{M}[m]$ and the reduced postinteraction detector state $\hat{\rho}'_D$. The third equality is the system picture involving the instrument $\mathcal{A}[m]$ acting on the initial system state $\hat{\rho}_S$. Since no further measurements are performed, the system picture can also be written as the last equality, which is the standard form for an expectation value of an effective system observable $\hat{A}_e[m]$ in the initial system state $\hat{\rho}_S$.

Due to the cyclic property of the trace, this effective observable can be understood as the action of the adjoint instrument $\mathcal{A}^*[m] = \sum_k m_k \mathcal{A}_k^*$ on the identity

$$\hat{A}_e[m] = \mathcal{A}^*[m](\hat{1}), \quad (10)$$

where the adjoint instrument is composed of adjoint quantum operations

$$\mathcal{A}_k^*(\cdot) = \langle \hat{U}^\dagger[(\cdot) \otimes |k\rangle \langle k|] \hat{U} \rangle_D = \sum_l \hat{M}_{k,l}^\dagger(\cdot) \hat{M}_{k,l}. \quad (11)$$

When acting on the identity, these adjoint operations induce a probability-operator measure (POM) [62]

$$\hat{P}_k = \mathcal{A}_k^*(\hat{1}) = \sum_l \hat{M}_{k,l}^\dagger \hat{M}_{k,l} \quad (12)$$

over the outcomes k of the detector such that $p_k = \langle \hat{P}_k \rangle_S$ are the measured detector probabilities. If all outcomes of the detector are accounted for, then the positive operators in the POM satisfy the normalization condition $\sum_k \hat{P}_k = \hat{1}$, making them a partition of unity. In terms of this POM, the effective system observable has the intuitive form

$$\hat{A}_e[m] = \sum_k m_k \hat{P}_k. \quad (13)$$

C. Measured observable

The operator $\hat{A}_e[m]$ in Eq. (13) is the precise *system* observable that is indirectly measured by the experimental procedure and chosen spectrum m . Due to the operator equality in Eq. (13), the detector faithfully measures this observable with any initial system state $\hat{\rho}_S$. Note, however, that the experimenter-assigned values m_k in the expansion of Eq. (10) are not generally the eigenvalues of $\hat{A}_e[m]$. Indeed, the number of detector outcomes k may in fact be drastically different than the number of eigenvalues of $\hat{A}_e[m]$.

When the POM \hat{P}_k consists of projection operators then it follows that $\hat{A}_e[m]$ and $\hat{M}[m]$ have the same spectrum and the equality of Eq. (10) reduces to the spectral expansion of $\hat{A}_e[m]$. However, when the POM is not projective, then the function m still constitutes a generalized spectrum for the measured observable $\hat{A}_e[m]$ that corresponds to the specific induced POM \hat{P}_k for the measurement. Such a generalized spectrum was dubbed a set of *contextual values* for the observable $\hat{A}_e[m]$ in previous work [44,52,53], since the experimentally relevant set of values for an indirectly measured observable will depend on the context of exactly how it is being measured.

III. ESTIMATION ERROR

Now suppose that we want to use an indirect measurement with instrument $\mathcal{A}[m]$ to estimate a particular system observable $\hat{A} = \sum_a A_a \hat{\Pi}_a$, with eigenvalues A_a and spectral projectors $\hat{\Pi}_a$. How do we quantify the error of such an estimation? To address this question, the inequalities of Eqs. (3)–(6) use the quantity

$$\epsilon_A^2 = \langle \hat{N}^2 \rangle, \quad (14)$$

which is the second moment of a joint *noise operator*

$$\hat{N} = \hat{U}^\dagger(\hat{1} \otimes \hat{M}[m]) \hat{U} - \hat{A} \otimes \hat{1} \quad (15)$$

under a specific initial joint state. Ozawa demonstrated [21,23,34] that $\epsilon_A = 0$ if and only if the estimation is “precise” or “noiseless,” and that this definition reduces to the classical notion of mean-squared error when the two terms of Eq. (15)

commute [11]. Hence, this definition seems like a natural choice for quantizing the classical notion of mean-squared error.

In our notation, such a “noiseless” estimation implies the operator equality $\hat{A} = \hat{A}_e[m]$ between the effectively measured observable and the desired system observable. Moreover, it implies that the induced POM \hat{P}_k must consist of the spectral projectors $\hat{\Pi}_a$ of \hat{A} , making it a projective measurement with the nonzero assigned values m_k equal to the eigenvalues A_a . In such a case, the number of nonzero m_k must match the number of eigenvalues indexed by a , so that the detector spectrum and the observable spectrum are essentially identical. Ozawa has argued that since the detector performs a projective measurement of the system observable in this case, then the system and detector are perfectly correlated and thus effectively “have the same value(s)” [59], which justifies his terminology of “noiseless” estimation.

To better understand the quantity ϵ_A , it is instructive to further dissect the noise operator \tilde{N} in terms of the system picture with the detector traced out. Observe that the first moment of the noise operator is

$$\delta_A = \langle \tilde{N} \rangle = \langle \hat{A}_e[m] - \hat{A} \rangle_S, \quad (16)$$

which measures the difference in the estimated mean from the target value. The second term involving the target value could be measured in a separate reference experiment, at least in principle.

In contrast, the second noise moment from Eq. (14) simplifies to a less intuitive expression in the system space

$$\epsilon_A^2 = \langle \hat{A}_e[m^2] \rangle + \hat{A}^2 - 2\hat{A}_e[m] * \hat{A} \rangle_S \quad (17)$$

that contains the symmetric Jordan product [introduced in Eq. (2)] between the measured observable $\hat{A}_e[m]$ and the target observable \hat{A} .

The first term of Eq. (17) is the second moment of the measured detector observable $\hat{M}[m]$, which is equivalent to the effective system observable $\hat{A}_e[m^2] = \sum_k m_k^2 \hat{P}_k$ obtained by squaring the spectrum m . Note that $\hat{A}_e[m]^2 \neq \hat{A}_e[m^2]$ unless the POM is projective.

The second term of Eq. (17) is the second moment of the target observable \hat{A} in the initial system state. As with the first moment in Eq. (16), this term could be measured in principle using a separate reference experiment.

The third term in Eq. (17), on the other hand, contains the Jordan product and does not obviously correspond to any measurable quantity pertaining to either the estimation experiment or a reference experiment. Hence, we have a conundrum: Though the squared noise operator \tilde{N}^2 is formally a Hermitian observable that generalizes the notion of classical mean-squared error, it appears to be constructed from quantities that are not operationally meaningful for the situation under consideration.

To resolve this conundrum, Ozawa [34] noted that it is possible to indirectly determine the problematic third term of Eq. (17) if one is able to prepare not just one, but *three* related system states: $\hat{\rho}_S$, $\hat{A}\hat{\rho}_S\hat{A}$, and $(\hat{1} + \hat{A})\hat{\rho}_S(\hat{1} + \hat{A})$. This indirect

procedure follows from the identity

$$2\langle \hat{A}_e[m] * \hat{A} \rangle_S = \langle (\hat{1} + \hat{A})\hat{A}_e[m](\hat{1} + \hat{A}) \rangle_S - \langle \hat{A}_e[m] \rangle_S - \langle \hat{A}\hat{A}_e[m]\hat{A} \rangle_S. \quad (18)$$

The additional operators \hat{A} and $(\hat{1} + \hat{A})$ that modify the state $\hat{\rho}$ in each term can be understood as the Kraus operators for an auxiliary preparation apparatus. Hence, all three states may be prepared without knowing $\hat{\rho}$ *a priori*, at least in principle, so each term in Eq. (18) can be measured in different reference experiments. This indirect procedure has been subsequently verified by Erhart *et al.* [35], Sulyok *et al.* [36], and Baek *et al.* [37].

While this indirect procedure clarifies that the quantity ϵ_A^2 can be experimentally determined, it leaves several remaining conundrums. First, it is not yet clear that ϵ_A^2 is still operationally meaningful for a single experiment, even if it can be indirectly determined in multiple experiments. Second, it is not clear exactly how ϵ_A^2 corresponds to a “mean-squared error” quantum mechanically, since that intuition was based on a classical analogy.

A. Quasiprobability interpretation

To address these remaining conundrums, Lund and Wise-man [38] noticed that the Jordan product of two observables can be given a restricted interpretation as a meaningful quantity in terms of quasiprobabilities. To see this, we also interpret the first two terms in Eq. (17) as Jordan products with an appropriate identity operator, and expand ϵ_A^2 directly in terms of the experimentally assigned values m_k and the eigenvalues A_a ,

$$\epsilon_A^2 = \sum_{k,a} (A_a - m_k)^2 \tilde{p}(a,k). \quad (19)$$

The distribution weighting this squared difference is a joint Terletsky-Margenau-Hill [39,40] *quasiprobability distribution*

$$\tilde{p}(k,a) = \langle \hat{\Pi}_a * \hat{P}_k \rangle_S, \quad (20)$$

which is the real part of the Dirac [63–65], or standard-ordering distribution [66]. It explicitly involves the symmetric Jordan product of the spectral projection operators $\hat{\Pi}_a$ of \hat{A} and the measured POM \hat{P}_k [67]. We use the tilde notation to indicate the quasiprobabilistic nature of the distribution, which can have negative values.

It is now easy to see that if $\tilde{p}(a,k)$ were a true joint probability distribution, then Eq. (19) would indeed compute the proper mean-squared deviation between the assigned values m_k of the estimation and the eigenvalues A_a of the target observable. However, $\tilde{p}(a,k)$ is a quasiprobability distribution that is not positive definite unless $\hat{A}_e[m]$ and \hat{A} commute with each other or the initial system state. This lack of positivity is a manifestation of the Bell-Kochen-Specker-Spekkens contextuality theorems [54–58]. Hence, Eq. (19) generally represents a mean-squared deviation only in a hypothetical, or *counterfactual* sense. It is the hypothetical deviation that *would* correspond to an estimation error *if* one could assign values to both observables \hat{A} and $\hat{A}_e[m]$ simultaneously even without measuring them; however, the contextuality theorems

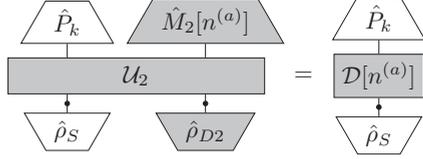


FIG. 3. Weak measurement used to approximate the quasiprobabilities $\tilde{p}(a,k)$ according to Eqs. (23) and (24), highlighted in gray. A weak interaction \mathcal{U}_2 couples a second detector state $\hat{\rho}_{D2}$ to the system, which is followed by a measurement of the detector observable $\hat{M}_2[n^{(a)}]$ with assigned values $n_\ell^{(a)}$. This joint procedure is equivalent to acting with a second quantum instrument $\mathcal{D}[n^{(a)}]$ on the system prior to measuring \hat{P}_k , with operations \mathcal{D}_ℓ corresponding to each outcome ℓ . This detector is tuned to measure a spectral projection $\mathcal{D}^*[n^{(a)}](\hat{1}) = \hat{\Pi}_a$ of \hat{A} when the \hat{P}_k measurement is ignored. However, subsequently measuring each k results in the averages $\sum_\ell n_\ell^{(a)} p(\ell,k) \approx \tilde{p}(a,k)$ that approximate a quasiprobability distribution when \mathcal{D}_ℓ is nearly the identity for all ℓ . Changing the assigned values $n_\ell^{(a)}$ retargets the detector for different spectral projectors $\hat{\Pi}_a$.

prohibit exactly such a hypothetical joint value assignment, even when the observables commute.

Nevertheless, these joint quasiprobabilities can still be approximately determined by introducing another detector into the experiment, as illustrated in Fig. 3. To see this, we use Bayes' rule to split the joint quasiprobabilities into a product

$$\tilde{p}(k,a) = \tilde{p}(a|k) p_k, \quad (21)$$

of the true detection probabilities $p_k = \langle \hat{P}_k \rangle_S$ and *conditional quasiprobabilities* [41]

$$\tilde{p}(a|k) = \text{Re} \frac{\text{Tr}_S[\hat{P}_k \hat{\Pi}_a \hat{\rho}_S]}{\text{Tr}_S[\hat{P}_k \hat{\rho}_S]}. \quad (22)$$

These conditional quasiprobabilities have the form of (real) *generalized weak values* [42–44,52,53,68–72] of the projection operators $\hat{\Pi}_a$ with “postselections” corresponding to the measured POM \hat{P}_k .

These weak values can then be approximately measured according to Fig. 3 by indirectly measuring the projection operators $\hat{\Pi}_a$ using a second coupled detector [38,45–47]. This second detector corresponds to an instrument $\mathcal{D}[n^{(a)}]$ with values $n_\ell^{(a)}$ assigned to its outcomes indexed by ℓ . When the original instrument $\mathcal{A}[m]$ is ignored, these values calibrate this new detector to measure a spectral projector of \hat{A} according to the identity $\hat{\Pi}_a = \mathcal{D}^*[n^{(a)}](\hat{1})$ analogous to Eq. (10). However, conditioning each outcome ℓ on a subsequently measured outcome k of the instrument $\mathcal{A}[m]$ instead produces the conditioned average

$$\sum_\ell n_\ell^{(a)} p(\ell|k) = \frac{\sum_\ell n_\ell^{(a)} p(\ell,k)}{\sum_\ell p(\ell,k)} \quad (23)$$

in terms of the measured joint probabilities $p(\ell,k) = \text{Tr}\{\mathcal{A}_k[\mathcal{D}_\ell(\hat{\rho}_S)]\} = \text{Tr}_S[\hat{P}_k \mathcal{D}_\ell(\hat{\rho}_S)] = \langle \mathcal{D}_\ell^*(\hat{P}_k) \rangle_S$. This conditioned average approximates a weak value equal to the quasiprobability $\tilde{p}(a|k)$ when the detector operations \mathcal{D}_ℓ are sufficiently close to the identity operation for every outcome ℓ [44,52,53].

We can also directly compute the *measured* distribution that approximates the quasiprobability distribution of Eq. (21) without the appeal to intermediary weak values. Since the measured p_k in Eq. (21) is $\sum_\ell p(\ell,k)$, which is the denominator of the conditioned average in Eq. (23), we immediately obtain

$$\sum_\ell n_\ell^{(a)} p(\ell,k) = \langle \mathcal{D}^*[n^{(a)}](\hat{P}_k) \rangle_S \approx \tilde{p}(a,k). \quad (24)$$

This distribution is what was determined in the experiments by Rozema *et al.* [45], Weston *et al.* [47], and Kaneda *et al.* [46].

Introducing such an auxiliary weak detector necessarily perturbs the initial system state differently for each of the outcomes ℓ , and thus modifies the experiment under consideration in a complicated way on average. However, within some error tolerance each detection probability p_k can be left approximately unaltered, and the joint distribution $\tilde{p}(a,k)$ can be approximately determined. This solution, however, raises the question of why the definition of intrinsic estimation error of a detector requires the introduction of a second detector with its own estimation error and resulting disturbance on the initial state. Determining the estimation error of that second detector would require a third detector, and so on.

B. Unbiased measurements

The interpretation of ϵ_A^2 as a mean-squared error averaged with quasiprobabilities can be avoided, however, in the special case of an unbiased measurement, where one demands the estimated mean to be faithful (i.e., $\delta_A = 0$) for any initial state. An operator equality $\hat{A}_e[m] = \hat{A}$ then follows from Eq. (16) and the positivity of $\hat{\rho}_S$ [44,52,53].

Due to this operator equality, Eq. (17) simplifies to

$$\epsilon_A^2 = \langle \hat{A}_e[m^2] - \hat{A}^2 \rangle_S = \sum_k m_k^2 p_k - \sum_a A_a^2 p_a, \quad (25)$$

which is now completely analogous to the first moment δ_A in Eq. (16). It is precisely the difference between the measured second moment using the raw detector values m_k and the ideal second moment of \hat{A} that would be measured with its eigenvalues A_a in a reference experiment. The concreteness of this expression occurs because the quasiprobabilities $\tilde{p}(a,k)$ in Eq. (19) become positive semidefinite when $\hat{A}_e[m]$ and \hat{A} commute.

The second noise moment thus quantifies the amplification of the signal spread due to the weakened correlation with an unbiased detector. Phrased in a different way, it quantifies the degree to which the assigned detector values have been amplified from the eigenvalues in order to compensate for the ambiguity of the measurement [44,52]. It is now easy to see why $\epsilon_A = 0$ if and only if the measurement is projective: Only in that case will unbiased detector values match the eigenvalues for the observable.

Importantly, however, ϵ_A does not indicate the quality of the estimation of the observable *distribution* that can be achieved using the experimental apparatus. Indeed, the operator equality $\hat{A}_e[m] = \hat{A}$ guarantees that the first moment can be precisely obtained given a sufficiently large statistical ensemble of measurements. Moreover, the same technique used to determine the spectral function m to produce this

equality can often be used to determine *different* spectral functions $m^{(n)}$ that satisfy other equalities $\hat{A}_e[m^{(n)}] = \hat{A}^n$ for all powers of \hat{A} [44,52]. In such a case, the same experimental data can be used to construct all higher moments of \hat{A} from the same measured set of probabilities, making the total experimental estimation of the distribution of \hat{A} essentially exact.

To emphasize this point, consider the following diagonal two-outcome POM for a qubit:

$$\hat{P}_1 = \begin{pmatrix} p_{11} & 0 \\ 0 & p_{12} \end{pmatrix}, \quad \hat{P}_2 = \begin{pmatrix} p_{21} & 0 \\ 0 & p_{22} \end{pmatrix}, \quad (26)$$

where $\hat{P}_1 + \hat{P}_2 = \hat{1}$, and the p_{ij} are positive probabilities such that $p_{i1} \neq p_{i2}$. Examples of how to implement a POM of this sort can be found in Refs. [46,47,73–79]. It is straightforward to show [52] that the POM in Eq. (26) can be used to uniquely construct any diagonal operator \hat{A} according to

$$\hat{A} = \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix} = m_1 \hat{P}_1 + m_2 \hat{P}_2, \quad (27a)$$

$$m_1 = \frac{p_{22} a_1 - p_{21} a_2}{p_{11} p_{22} - p_{21} p_{12}}, \quad (27b)$$

$$m_2 = -\frac{p_{12} a_1 - p_{11} a_2}{p_{11} p_{22} - p_{21} p_{12}}. \quad (27c)$$

Evidently the higher powers of \hat{A} are special cases of this solution that are obtained through the simple replacement $a_j \rightarrow (a_j)^n$ in the values $m_k^{(n)}$ for \hat{A}^n . Thus, all the moments of \hat{A} can be determined in *one experimental run* of the detector by assigning appropriate values to each detector outcome in *postprocessing*. Notably, this method to determine the observable moments using a noisy signal is essentially classical in character since Eqs. (26) and (27) involve diagonal operators [52].

It follows that one can construct $\hat{A}^2 = \hat{A}_e[m^{(2)}]$ with this technique, so one can compute $\langle \hat{A}^2 \rangle_S$ directly from the same data used to compute $\langle \hat{A}_e[m] \rangle_S$. Therefore, the second noise moment of Eq. (25) can be directly determined with no additional experimental work as the simple expression

$$\epsilon_A^2 = \sum_k [(m_k)^2 - m_k^{(2)}] p_k. \quad (28)$$

Hence, no reference experiments, quasiprobability arguments, or auxiliary detectors should be needed to directly determine this quantity in the laboratory, provided that one is using an unbiased detector.

This demonstration also implies that the quantity ϵ_A in Eq. (14) can be nonzero even when all moments of the measured observable can be faithfully determined by the same experimental data. As such, it does not quantify the estimation error for the distribution of the observable. A better name for ϵ_A is the dispersion of the estimation of the *mean* of the observable, as also pointed out by Hall [22]. For an experimenter, a large dispersion corresponds directly to the need for an increased number of measurement realizations to produce good statistical results. The more ambiguous (i.e., weak, or noisy) the measurement is, the larger the detector values must be, so the number of realizations needed to

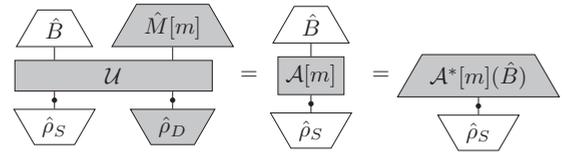


FIG. 4. Sequential measurements. After the indirect measurement illustrated in Fig. 2 and highlighted in gray, another system observable \hat{B} is measured. The procedure produces correlated pairs of measurement outcomes (m_k, B_b) . The quantum instrument $\mathcal{A}[m]$ is needed to fully describe these correlations in the reduced system space. The effective system observable $\hat{A}_e[m]$ from Fig. 2 reappears from the sequential observable $\hat{A}^*[m](\hat{B})$ only when the correlations from the measurement of \hat{B} are ignored.

obtain the same statistical error for the mean will be larger. Nevertheless, for a sufficiently large number of realizations even noisy measurements can be made statistically precise.

The interpretation of ϵ_A as a generic estimation error (as opposed to the dispersion of the mean) thus hinges crucially upon its decomposition into a mean-squared error for the *individual measurement realizations*. Since it does not indicate the estimation error for the full distribution of \hat{A} , one can only argue that it provides a sensible notion of the average estimation error for each realization. However, to adopt this point of view is to assert that the measured observable has a definite (correct) value prior to each measurement to which the (incorrect) detector result can be compared. While this assertion is unproblematic for classical systems with definite values, it is a nontrivial assertion for a quantum-mechanical system. In contrast, interpreting ϵ_A as the dispersion in the estimation of the mean of the observable does not demand such a controversial statement.

IV. MEASUREMENT DISTURBANCE

Now suppose we wish to measure a second observable $\hat{B} = \sum_b B_b \hat{\Pi}_b$ with eigenvalues B_b and spectral projections $\hat{\Pi}_b$ after the indirect measurement of $\mathcal{A}[m]$, as illustrated in Fig. 4. To what degree has the act of measuring $\mathcal{A}[m]$ disturbed the subsequent measurement of \hat{B} ? To address this question, the inequalities of Eqs. (3) and (6) use the quantity

$$\eta_B^2 = \langle \hat{D}^2 \rangle, \quad (29)$$

which is analogous to Eq. (14), and is the second moment of a joint *difference* operator

$$\hat{D} = \hat{U}^\dagger (\hat{B} \otimes \hat{1}) \hat{U} - \hat{B} \otimes \hat{1}, \quad (30)$$

between the original observable \hat{B} and the joint observable modified by the unitary interaction in the Heisenberg picture. Ozawa demonstrated [21,23,34] that $\eta_B = 0$ if and only if the estimation of \hat{A} does not affect the subsequent measurement of \hat{B} .

Again, it is instructive to dissect this definition in terms of the system picture with the detector traced out. First note that the detector observable $\hat{M}[m]$ with spectral function m has been replaced by the identity $\hat{M}[1] = \hat{1}$ in the definition of Eq. (29). Setting the values m_k to 1 in this manner marginalizes over all the detector outcomes, which performs a *nonselective*

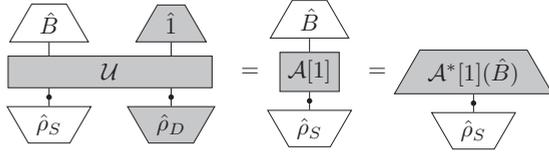


FIG. 5. Disturbance induced by a nonselective measurement. After the nonselective measurement highlighted in gray, another system observable \hat{B} is measured. Unlike in Fig. 4, this procedure ignores the correlations between specific pairs of measurement outcomes, so will measure an average perturbed observable $\hat{B}' = \mathcal{A}^*[1](\hat{B})$.

measurement using the entire apparatus \mathcal{A} . This replacement implies that any information about the correlations between pairs (m_k, B_b) of sequentially measured outcomes is being discarded by this procedure. Only the net effect of averaging over *all* outcomes k is described by Eq. (29), which is better illustrated by Fig. 5.

We can write the first moment of the difference operator of Eq. (30) in several ways:

$$\delta_B = \langle \hat{D} \rangle = \text{Tr}_S[\hat{B}(\hat{\rho}'_S - \hat{\rho}_S)] = \langle \hat{B}' - \hat{B} \rangle_S, \quad (31)$$

where

$$\hat{\rho}'_S = \mathcal{A}[1](\hat{\rho}_S) = \sum_k \mathcal{A}_k(\hat{\rho}_S) = \sum_{k,l} \hat{M}_{k,l} \hat{\rho}_S \hat{M}_{k,l}^\dagger \quad (32)$$

is the reduced postinteraction state of the system, and

$$\hat{B}' = \mathcal{A}^*[1](\hat{B}) = \sum_k \mathcal{A}_k^*(\hat{B}) = \sum_{k,l} \hat{M}_{k,l}^\dagger \hat{B} \hat{M}_{k,l} \quad (33)$$

is the Heisenberg operator that has been perturbed by the nonselective measurement. Thus, the difference operator in Eq. (31) provides information about how the mean of \hat{B} changes due to the evolution induced by the nonselective measurement.

The squared disturbance in Eq. (29), on the other hand, reduces to the expression

$$\eta_B^2 = \langle (\hat{B}^2)' + \hat{B}^2 - 2\hat{B}' * \hat{B} \rangle_S \quad (34)$$

on the system space, where

$$(\hat{B}^2)' = \mathcal{A}^*[1](\hat{B}^2) = \sum_{k,l} \hat{M}_{k,l}^\dagger \hat{B}^2 \hat{M}_{k,l} \quad (35)$$

is the squared Heisenberg operator perturbed by the nonselective measurement. Analogously to Eq. (17), the expression of Eq. (34) contains a symmetric Jordan product between the original observable \hat{B} and the perturbed observable \hat{B}' , so is challenging to interpret in an experimentally meaningful way. One can still indirectly determine the Jordan product term by exploiting the same identity used in Eq. (18) of Sec. III, however.

A. Quasiprobability interpretation

As with the dispersion of Eq. (17), Lund and Wiseman [38] observed that one can obtain an operational meaning for Eq. (34) by expanding the Jordan product into a joint Terletsky-Margenau-Hill quasiprobability distribution. To do so, we interpret the first two terms of Eq. (34) as Jordan

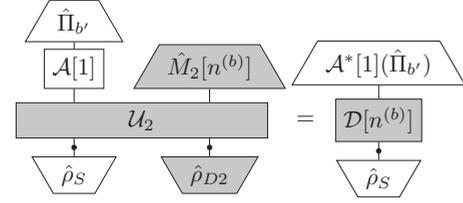


FIG. 6. Weak measurement used to approximate the quasiprobabilities $\tilde{p}(b, b')$ in Eq. (37), highlighted in gray. A second quantum instrument $\mathcal{D}[n^{(b)}]$ is coupled to the system prior to the nonselective measurement $\mathcal{A}[1]$. This detector is tuned to measure a spectral projection $\mathcal{D}^*[n^{(b)}](\hat{I}) = \hat{\Pi}_b$ of \hat{B} when all subsequent measurements are ignored. However, measuring each b' after $\mathcal{A}[1]$ results in the averages $\sum_{\ell,k} n_\ell^{(b)} p(\ell, k, b') \approx \tilde{p}(b, b')$ that approximate a quasiprobability distribution when \mathcal{D}_ℓ is nearly the identity for all ℓ . Changing the assigned values $n_\ell^{(b)}$ retargets the detector for different spectral projectors $\hat{\Pi}_b$.

products with an appropriate identity operator, and expand η_B^2 directly in terms of the experimentally assigned values $B_{b'}$, the eigenvalues B_b ,

$$\eta_B^2 = \sum_{b,b'} (B_{b'} - B_b)^2 \tilde{p}(b', b), \quad (36)$$

and the joint *quasiprobability distribution*

$$\tilde{p}(b', b) = \langle \hat{Q}_{b'} * \hat{\Pi}_b \rangle_S \quad (37)$$

that involves the Jordan product of the spectral projection operators of $\hat{B} = \sum_b B_b \hat{\Pi}_b$ and the perturbed POM $\hat{Q}_{b'} = \mathcal{A}^*[1](\hat{\Pi}_{b'})$ that constructs the perturbed observable $\hat{B}' = \sum_{b'} B_{b'} \hat{Q}_{b'}$ actually measured.

As before, this joint quasiprobability distribution is not positive definite unless \hat{B}' and \hat{B} commute with each other or the initial system state. Nevertheless, it can be approximately measured by probing the system weakly prior to the nonselective interaction of the apparatus \mathcal{A} [38,45,47], as illustrated in Fig. 6. This procedure is entirely analogous to the one discussed in Fig. 3 and Sec. III A.

Since determining a joint *quasiprobability distribution* is necessary in order to interpret Eq. (34) as a mean-squared error caused by disturbance, the quantity η_B^2 does not generally pertain to any concrete notion of the disturbance inflicted upon \hat{B} by the apparatus \mathcal{A} . Instead, η_B^2 pertains to a hypothetical, or *counterfactual*, notion of disturbance. That is, the observable \hat{B} is implicitly assigned a *definite* value B_b prior to the interaction for each realization. That value is then disturbed to a different but equally definite value $B_{b'}$ with some transition quasiprobability $\tilde{p}(b', b)$ that averages over the effects of all intermediate k . The nonpositivity of the quasiprobability distribution indicates the questionable nature of this assumption that is incompatible with the Bell-Kochen-Specker-Spekkens contextuality theorems [54–58].

B. Quantum nondemolition measurements

If one demands that $\delta_B = 0$ for any initial state, then another operator equality $\hat{B}' = \hat{B}$ follows from Eq. (31) and the positivity of $\hat{\rho}_S$. This equality is satisfied when $(\hat{B} \otimes \hat{I})$ commutes with \hat{U} , or, equivalently, when \hat{B} commutes with

$\hat{M}_{k,l}$ for all k,l . Notably, these commutation conditions are precisely the criteria for \mathcal{A} to be a *quantum nondemolition* (QND) measurement with respect to \hat{B} [80].

For such a QND measurement, the perturbed POM in Eq. (37) $\hat{Q}_{b'} \rightarrow \hat{\Pi}_{b'}$ reduces to the spectral projections of \hat{B} (up to relabeling of indices), the joint quasiprobability distribution becomes a true diagonal joint probability operator $\hat{p}(b',b) \rightarrow p_b \delta_{b,b'}$, and the disturbance trivially vanishes: $\eta_B = 0$.

C. Lindblad decoherence

The perturbation to \hat{B} can also be understood in terms of the induced *Lindblad decoherence* stemming from the flow of system information to the discarded detector. We can emphasize this connection by expanding the perturbed operator \hat{B}' using the following identity for each k :

$$\sum_l \hat{M}_{k,l}^\dagger \hat{B} \hat{M}_{k,l} = \hat{P}_k * \hat{B} + \mathcal{L}_k(\hat{B}), \quad (38)$$

where the probability operator $\hat{P}_k = \sum_l \hat{M}_{k,l}^\dagger \hat{M}_{k,l}$ appears in a Jordan product, while the remainder $\mathcal{L}_k = \sum_l \mathcal{L}[\hat{M}_{k,l}^\dagger]$ is composed of *Lindblad operations* [81]

$$\mathcal{L}[\hat{M}^\dagger](\hat{B}) = -\frac{1}{2}(\hat{M}^\dagger[\hat{M}, \hat{B}] - [\hat{M}^\dagger, \hat{B}]\hat{M}). \quad (39)$$

These operations were introduced in the study of decoherence for open quantum systems [82]. In the present context the Lindblad operation indicates the average perturbation to \hat{B} that is induced by a particular measurement operator \hat{M} .

After summing the identity of Eq. (38) over all k , we find the intuitive relation

$$\hat{B}' - \hat{B} = \sum_k \mathcal{L}_k(\hat{B}) \quad (40)$$

between \hat{B} and its perturbation \hat{B}' . The difference in the measured mean δ_B from Eq. (31) depends solely on this difference, so will be governed by the *net* induced perturbation from the nonselective measurement.

The same procedure can be applied to the second moment (and indeed all moments) to obtain the relation

$$(\hat{B}^2)' - \hat{B}^2 = \sum_k \mathcal{L}_k(\hat{B}^2). \quad (41)$$

We can thus insert the equalities of Eqs. (40) and (41) into Eq. (34) to obtain

$$\eta_B^2 = \sum_k \langle \mathcal{L}_k(\hat{B}^2) - 2\hat{B} * \mathcal{L}_k(\hat{B}) \rangle_S. \quad (42)$$

Hence, only the Lindblad perturbation terms contribute to the squared disturbance η_B^2 . However, while both Eqs. (40) and (41) indicate measurable aspects of the disturbance to the moments of the distribution of \hat{B} , the constructed quantity in Eq. (42) still has an additional hypothetical character due to the remaining Jordan product. In this sense, the manner in which the quantity η_B quantifies the average perturbation to \hat{B} is unnatural from a distributional perspective.

V. AN ALTERNATIVE PERSPECTIVE

Heisenberg's original discussion in Ref. [1], on which the inequalities of Eqs. (3)–(6) are ostensibly based, was

concerned with a different sort of estimation error and disturbance than we have been analyzing. Rather than hypothetically tracking individual observable values being measured with all outcomes of an apparatus, Heisenberg considered how individual particles were actively affected by a *single outcome* of the apparatus on average. Depending on the ambiguity inherent to each apparatus outcome (e.g., a microscope with finite resolution for each measured interval), the future spreads of conjugate quantities would be altered in complementary ways on average. This idea of average disturbance and estimation error corresponding to single apparatus events is not adequately captured by the definitions of ϵ_A and η_B , which rely on information obtained from all outcomes of the apparatus on average. This point has also been made by Werner [26], Busch *et al.* [8,28,32,33], Watanabe *et al.* [29,30], and Hofmann [25,31].

As such, we feel that the inequalities of Eqs. (3)–(6) do not capture the spirit of Heisenberg's original uncertainty discussion. Instead, they are independently interesting and qualitatively different preparation-dependent measures of the dispersion of the estimation of the mean and the average quasiperturbation of individual eigenvalues. We are thus led to consider alternative definitions of estimation error and disturbance that depend only on the individual instrument outcomes themselves. This line of thought leads directly to several inequalities that have the same form as the Heisenberg relation in Eq. (1), and which generalize the results in Ref. [25].

A. Estimation error

Since Heisenberg's arguments pertain directly to how individual outcomes of the measurement instrument affect any prepared particle on average, we examine the quantum instrument $\mathcal{A}[m]$ that is used to make the estimation without any reference to an initial system state. Each outcome k of this instrument produces a POM element \hat{P}_k according to Eq. (7). As discussed in Refs. [53,83,84], normalizing this POM element produces a *retrodictive state* associated with the outcome k

$$\check{\rho}_k = \frac{\hat{P}_k}{\text{Tr}_S[\hat{P}_k]} = \frac{\sum_l \hat{M}_{k,l}^\dagger \hat{M}_{k,l}}{\sum_l \text{Tr}_S[\hat{M}_{k,l}^\dagger \hat{M}_{k,l}]}, \quad (43)$$

where we introduce an inverted hat to indicate the retrodictive nature of the state.

A natural notion of estimation error can then be defined as the *retrodictive* standard deviation, $\epsilon_{A,k}$, which is the uncertainty in \hat{A} that can be retroactively inferred on average after obtaining the single outcome k on the detector. This is the best average uncertainty that one can infer with no prior information about the particle, and may be understood as the *resolution* of the detector outcome k . It is defined from the variance with respect to the retrodictive state for the outcome k ,

$$\epsilon_{A,k}^2 = \sum_a (A_a - \langle \hat{A} \rangle_k)^2 p(a|k) = \langle \hat{A}^2 \rangle_k - \langle \hat{A} \rangle_k^2, \quad (44)$$

where $\langle \cdot \rangle_k = \text{Tr}_S[(\cdot)\check{\rho}_k]$ is the retrodictive expectation value, and $p(a|k) = \langle \hat{\Pi}_a \rangle_k$ is the retrodictive probability of a given the outcome k [53]. This quantity can be directly measured as illustrated in Fig. 7 by comparing specific preparations of \hat{A}

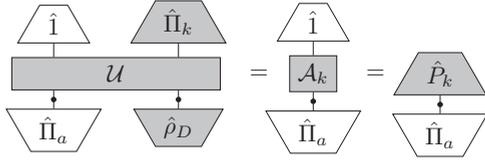


FIG. 7. Operational procedure for determining the estimation error $\epsilon_{A,k}$. For every preparation $\hat{\Pi}_a$ in the spectral basis of \hat{A} made with probability $p(a)$, the system is prepared in the initial state $\hat{\Pi}_a$ and subsequently conditioned on the successful measurement of a particular outcome k , which produces the measurable probabilities $p(k|a)$. Applying Bayes' rule yields the retrodictive probabilities $p(a|k) = p(k|a)p(a)/\sum_a p(k|a)p(a)$. For a uniform preparation with constant $p(a)$, this procedure produces $p(a|k) = \text{Tr}_S[\hat{\Pi}_a \hat{\rho}_k]$, which effectively normalizes the POM element \hat{P}_k to produce the retrodictive state $\hat{\rho}_k$ in Eq. (43) that averages each possible preparation $\hat{\Pi}_a$. These probabilities define $\epsilon_{A,k}$ according to Eq. (44). The assumption of uniform $p(a)$ indicates that a single independent event k cannot be assumed to correspond to any particular source preparation *a priori*. The resulting probabilities $p(a|k)$ are thus the best retrodictive inference about the preparation that one can make using only information about a single isolated detector outcome k .

with resulting outcomes k . Moreover, this quantity is a property of the specific instrument outcome k without any reference to any particular preparation [85].

As an example, consider a common qubit POM $\hat{P}_\pm = [\hat{1} \pm \cos \theta \hat{\sigma}_3]/2$ parametrized by an angle θ . This POM can be used to estimate the Pauli operator $\hat{\sigma}_3$. When $\theta = n\pi$ with integer n then the POM is projective and the estimation error for each outcome should be zero. Conversely, for $\theta = (n + 1/2)\pi$ each POM element becomes the identity operator $\hat{P}_\pm \rightarrow \hat{1}/2$, so no information may be inferred about $\hat{\sigma}_3$; therefore, the estimation error for each outcome should be maximal. Computing each estimation error in Eq. (44) yields the constant $\epsilon_{\sigma_3,\pm} = |\sin \theta|$, since the outcomes are symmetric. This average error describes the resolution of each outcome and has the correct θ dependence. Note that there is no *a priori* fixed correspondence between the outcomes \pm of the detector and the eigenvalues of $\hat{\sigma}_3$. As such, $\epsilon_{\sigma_3,\pm}$ automatically chooses the optimal correspondence that has a maximum error of 1 (as opposed to 2 if the outcomes were permitted to correspond incorrectly to the eigenvalues in the projective case).

For comparison, the dispersion ϵ_{σ_3} used by Ozawa and others in Eq. (14) depends not only on the entire POM, but also on the specific values m_\pm that are assigned to each detector outcome. Since ϵ_{σ_3} depends on this choice, it does not have a value that is intrinsic to the apparatus itself. Assigning the values $m_\pm = \pm 1/\cos \theta$ to the detector makes the estimation unbiased, meaning $\sum_\pm m_\pm \hat{P}_\pm = \hat{\sigma}_3$; this is a special case of Eq. (27). If we pick these values, then Eq. (25) immediately implies that $\epsilon_{\sigma_3}^2 = \sec^2 \theta - \langle \hat{\sigma}_3 \rangle_S^2 \geq \sec^2 \theta - 1 = \tan^2 \theta$. The second term is bounded by 1 and depends on the preparation state, while the first term is state agnostic and diverges as $\theta \rightarrow (n + 1/2)\pi$. The size of this dispersion is related to the number of realizations that are required to statistically determine the mean $\langle \hat{\sigma}_3 \rangle_S$ to an acceptable precision. However, it does not directly indicate the resolution of each apparatus outcome, in contrast to our alternative definition in Eq. (44).

Returning to the retrodictive estimation error, since Eq. (44) has the form of a standard deviation, the Weyl-Robertson inequality in Eq. (1) immediately implies the following inequality that generalizes one also derived by Hofmann [25],

$$\epsilon_{A,k} \epsilon_{B,k} \geq \left| \frac{1}{2i} \langle [\hat{A}, \hat{B}] \rangle_k \right|, \quad (45)$$

who considered outcomes with only a single measurement operator. The quantity $\epsilon_{B,k}$ is defined identically to $\epsilon_{A,k}$ in Eq. (44), but with the substitution of the operator \hat{B} . The bound involves the retrodictive expectation of the commutator between the two operators.

This inequality is a form of *estimation complementarity* that holds specifically for a single outcome k and agrees with the standard form of the Heisenberg uncertainty inequality. It indicates that a single apparatus outcome k cannot simultaneously estimate the values of two incompatible observables on average beyond a certain precision limited by an uncertainty bound. As such, it corresponds to the second statement made in the introduction regarding complementarity, albeit with the important replacement of the preparation state with the retrodictive state corresponding to a specific instrument outcome k . Indeed, note that for the canonical variables \hat{x} and \hat{p} considered by Heisenberg, this bound correctly reduces to the state-independent quantity $\hbar/2$. Moreover, this inequality can also be trivially improved via the Cauchy-Schwartz form of Eq. (2).

B. Measurement disturbance

To describe an operational notion of disturbance, we need to consider correlations between measurable events prior and posterior to an intermediate outcome k . As discussed in Ref. [53], conditioning on an intermediate k will normalize the corresponding transformation \mathcal{A}_k in Eq. (7), which produces an *interdictive state*

$$\tilde{\rho}_k(\cdot) = \frac{\mathcal{A}_k(\cdot)}{\text{Tr}_S[\mathcal{A}_k(\hat{1})]} = \frac{\sum_l \hat{M}_{k,l}(\cdot) \hat{M}_{k,l}^\dagger}{\sum_l \text{Tr}_S[\hat{M}_{k,l} \hat{M}_{k,l}^\dagger]} \quad (46)$$

that encodes all transformation information about that single detector outcome while leaving the choice of prior and posterior measurements unspecified. This interdictive state is related to the retrodictive state for k according to $\tilde{\rho}_k = \hat{\rho}_k^*(\hat{1})$. We use the tilde notation here to indicate that this exotic type of state is a *transformation* and not simply an operator.

Using the interdictive state, we can naturally define an operational notion of disturbance as a mean-squared deviation between precise preparations and postselections of an observable \hat{B} that bracket a specific outcome k of the instrument

$$\eta_{B,k}^2 = \sum_{b,b'} (B_b - B_{b'})^2 p(b,b'|k). \quad (47)$$

Unlike the quasiprobability distribution in Eq. (36), the joint distribution

$$p(b,b'|k) = \text{Tr}_S[\hat{\Pi}_{b'} \tilde{\rho}_k(\hat{\Pi}_b)] = \frac{\sum_l |\langle b' | \hat{M}_{k,l} | b \rangle|^2}{\sum_l \text{Tr}_S[\hat{M}_{k,l} \hat{M}_{k,l}^\dagger]} \quad (48)$$

is a true probability distribution, and directly corresponds to the experimental method illustrated in Fig. 8 for determining

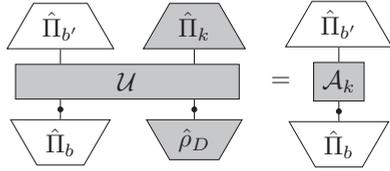


FIG. 8. Operational procedure for determining the measurement disturbance $\eta_{B,k}$. For each projection $\hat{\Pi}_b$ in the spectral basis of \hat{B} , the system is prepared in the initial state $\hat{\Pi}_b$ with probability $p(b)$, the detector outcome k is measured, then the outcome $\hat{\Pi}_{b'}$ is measured, producing the measurable probabilities $p(k, b'|b)$. Conditioning on the successful measurement of a particular intermediate outcome k produces the interdictive joint probabilities according to Bayes’ rule $p(b, b'|k) = p(k, b'|b)p(b) / \sum_{b, b'} p(k, b'|b)p(b)$. When the preparation $p(b)$ is uniform, this yields $p(b, b'|k) = \text{Tr}_S[\hat{\Pi}_{b'} \tilde{\rho}_k(\hat{\Pi}_b)]$, which effectively normalizes the operation \mathcal{A}_k to produce the interdictive state operation $\tilde{\rho}_k$ in Eq. (46) that correlates each preparation $\hat{\Pi}_b$ and postselection $\hat{\Pi}_{b'}$. These probabilities define $\eta_{B,k}$ according to Eq. (47). As with the procedure in Fig. 7, the assumption of uniform $p(b)$ indicates the lack of *a priori* bias that can be inferred by a single isolated outcome k .

how each outcome k disturbs definite preparations of \hat{B} on average. This definition of disturbance produces an intrinsic property of the instrument that depends only on the outcome k and the chosen observable \hat{B} .

Upon insertion of the Lindblad identity of Eq. (38), we can rewrite Eq. (47) as

$$\eta_{B,k}^2 = \sum_{b, b'} (B_b - B_{b'})^2 \frac{\langle b | \mathcal{L}_k(\hat{\Pi}_{b'}) | b \rangle}{\sum_l \text{Tr}_S[\hat{M}_{k,l} \hat{M}_{k,l}^\dagger]}. \quad (49)$$

This form demonstrates that only the Lindblad perturbation $\mathcal{L}_k = \sum_l \mathcal{L}[\hat{M}_{k,l}^\dagger]$ corresponding to the outcome k will contribute to the mean-squared deviation $\tilde{\eta}_{B,k}^2$ in a natural way, in contrast to Eq. (42).

Our definition of disturbance in terms of an interdictive state reproduces and generalizes the one used in Ref. [25] for an outcome with a single measurement operator. Hence, we can similarly define a restricted disturbance for each posterior b' ,

$$\begin{aligned} \eta_{B,k,b'}^2 &= \sum_b (B_b - B_{b'})^2 p(b|k, b') \\ &= \epsilon_{B,k,b'}^2 + (B_{b'} - \langle \hat{B} \rangle_{k,b'})^2, \end{aligned} \quad (50)$$

where $\langle \hat{B} \rangle_{k,b'} = \text{Tr}_S[\tilde{\rho}_{k,b'} \hat{B}]$ is the retrodictive mean under the state $\tilde{\rho}_{k,b'} = \tilde{\rho}_k^*(\hat{\Pi}_{b'}) / p(b'|k)$ conditioned on both k and b' . For each b' we then obtain the generalization to the second Hofmann inequality

$$\epsilon_{A,k,b'} \eta_{B,k,b'} \geq \epsilon_{A,k,b'} \epsilon_{B,k,b'}, \quad (51)$$

which, combined with Eq. (45), produces an estimation-disturbance inequality for each outcome pair (k, b') that agrees with the standard form of the Heisenberg uncertainty relation.

Formally averaging Eq. (51) over every b' with the probability distribution $p(b'|k) = \text{Tr}_S[\tilde{\rho}_k^*(\hat{\Pi}_{b'})]$ of obtaining b' given k produces the generalization of the third Hofmann

inequality [25]

$$\epsilon_{A,k} \eta_{B,k} \geq \left| \frac{1}{2i} \langle [\hat{A}, \hat{B}] \rangle_k \right| \quad (52)$$

that involves the full disturbance in Eq. (47) and the retrodictive state for k in a similar manner to the inequality of Eq. (45). The averaging step is a formal technique to derive the form of Eq. (52), which holds generally. Moreover, just as with Eq. (45), the inequality of Eq. (52) may be further improved to match the form of Eq. (2) if desired.

The inequality of Eq. (52) corresponds to the third statement made in the introduction associated with the Heisenberg uncertainty relation that pertains to the error-disturbance trade-off intrinsic to a specific measurement outcome k . As with the complementarity inequality, substituting the canonical operators \hat{x} and \hat{p} considered by Heisenberg produces the traditional state-independent bound of $\hbar/2$.

C. Discussion

To obtain results that are inconsistent with the inequalities of Eqs. (45) and (52), an experimenter must already know additional information about the initial state and combine that prior information with the new information obtained from the measurement. This requirement of possessing information from multiple points in time has been recently emphasized by Rozema *et al.* [86], who defend the inequalities of Eqs. (3)–(6) as explicitly including the complete information about the initial state. Such information is typically obtained from performing state tomography over an ensemble of different measurements made in separate experiments using the same preparation procedure. This sort of estimation using general priors from multiple points in time necessarily applies to ensembles of measurements made by the detector, as we emphasized in the first half of this paper.

Heisenberg, however, was concerned with what one could infer on average from a single event of a detector at a local point in time, which is a qualitatively different scenario. Indeed, according to Heisenberg [3, p. 20], his principle

“...states that every subsequent observation of the position will alter the momentum by an unknowable and undeterminable amount such that *after* carrying out the experiment *our knowledge* of the electronic motion is restricted by the uncertainty relation.” (emphasis added)

Indeed, any postmeasurement predictive state automatically satisfies the inequality (1) concerning the intrinsic spreads of future measurements. He goes on to clarify that [3, p. 20]

“...the uncertainty relation does not refer to the past; if the velocity of the electron is at first known and the position then exactly measured, the position for times previous to the measurement may be calculated. Then for these past times $\Delta p \Delta q$ is smaller than the usual limiting value, but this knowledge of the past is of a *purely speculative character*, since it can never...be subjected to experimental verification. It is a matter of personal belief whether such a calculation concerning the past history of the electron can be ascribed any physical reality or not.” (emphasis added)

Evidently, Heisenberg thinks it is inappropriate to make an inference by combining the average knowledge about potential

events at a prior point in time (i.e., the initial state) with detector events at a later point in time. That combined knowledge can only be of a “purely speculative character,” and thus lead to *counterfactual* characterizations of individual particle event chains, as we have pointed out with the definitions of error and disturbance used in the inequalities of Eqs. (3)–(6). This observation does not imply that these speculations about ensembles of past-particle histories are not interesting or important in their own right (e.g., [87]), but it does challenge their interpretation as simple corrections to the uncertainty principle of Heisenberg.

Using only information that pertains to a single detector event at a particular time, on the other hand, does satisfy the spirit of Heisenberg’s discussion of uncertainty. As we have shown, making this restriction immediately produces the inequalities Eqs. (45) and (52) for that isolated detector outcome, which complement the traditional Weyl-Robertson inequality in Eq. (1). Together, these inequalities correspond to all three interpretations of Heisenberg’s uncertainty principle outlined in the Introduction, as it applies to each outcome of a detecting apparatus. We also emphasize that other apparatus-intrinsic definitions of error and disturbance produce similar results [29,30,32,33], though we do not explore them here.

VI. CONCLUSION

In this work we have revisited the formal operator definitions of the mean-squared estimation error ϵ^2 and disturbance η^2 used in recent inequalities derived by Ozawa and others that ostensibly generalize the Heisenberg uncertainty relation. These important inequalities have recently been tested in *tour de force* experiments. Nevertheless, by analyzing the formal definitions in the system space, we have shown that they have unsatisfactory features that seem to deviate from the original discussion of Heisenberg.

Underlying the interpretation of the definitions of ϵ and η as mean-squared deviations is the hidden assumption that observables have definite but unknown value assignments even when they are not measured. This assumption is inconsistent with the Bell-Kochen-Specker-Spekkens contextuality theorems. The nonpositivity of the quasiprobability distributions required to average the deviations indicates the questionable nature of the hidden assumption. Moreover, these definitions pertain to inferences made from all outcomes of an apparatus on average, whereas Heisenberg considered what could be inferred from a single apparatus outcome on average.

For unbiased estimations, we clarified that the quantity ϵ can be given a concrete interpretation as the dispersion in the estimation of the mean. Furthermore, we demonstrated how this quantity (and every observable moment) can be directly determined with the same experimental data used to estimate the mean, even when using nonprojective measurements. As such, ϵ does not quantify an estimation error for the full observable distribution; thus, its interpretation as an estimation error rests squarely upon its questionable decomposition into an average error of the individual measurement realizations themselves.

To reconnect with Heisenberg’s original discussion, we considered alternative definitions of the estimation error and disturbance that focus on what one can infer on average from a *single outcome* of an experimental apparatus. By making the definition preparation agnostic, these alternative definitions describe properties intrinsic to the apparatus outcome itself. The estimation error ϵ_k of a detector outcome k indicates its intrinsic *resolution*, and has the form of a standard deviation with respect to the *retrodictive state* of the outcome. This definition produces a complementarity inequality for each k . These inequalities have the traditional form of the uncertainty relation, but substitute the retrodictive states for particular outcomes in place of the preparation state.

The analogous operational definition of disturbance η_k for the outcome k has the form of a root-mean-squared deviation between prior and posterior measurements conditioned on the intermediate outcome k . This correlation can be written in terms of the recently introduced *interdictive state* for the outcome. This definition produces an error-disturbance inequality for each outcome k . These inequalities also have the traditional form of the uncertainty relation, albeit with a similar substitution of the retrodictive states of each outcome.

Therefore, the traditional Heisenberg uncertainty relation has been vindicated for single apparatus outcomes, even when applied to complementarity and error-disturbance arguments.

ACKNOWLEDGMENTS

J.D. would like to thank Curtis J. Broadbent, Adam Miranowicz, Abraham G. Kofman, Mark J. Everitt, and Eyob A. Sete for stimulating and enlightening discussions. This work was partially supported by the ARO, RIKEN iTHES Project, MURI Center for Dynamic Magneto-Optics, JSPS-RFBR Contract No. 12-02-92100, Grant-in-Aid for Scientific Research (S), MEXT Kakenhi on Quantum Cybernetics, and the JSPS via its FIRST program.

[1] W. Heisenberg, *Z. Phys.* **43**, 172 (1927).

[2] E. H. Kennard, *Z. Phys.* **44**, 326 (1927).

[3] W. Heisenberg, *The Physical Principles of the Quantum Theory* (University of Chicago Press, Chicago, 1930) (reprinted by Dover, New York, 1949, 1967).

[4] H. Weyl, *Gruppentheorie Und Quantenmechanik* (Hirzel, Leipzig, 1928); Hermann Weyl, *The Theory of Groups and Quantum Mechanics* (Dover, New York, 1950).

[5] H. P. Robertson, *Phys. Rev.* **34**, 163 (1929).

[6] E. Schrödinger, *Sitzungsber. Preuß. Akad. Wiss. Berlin (Math. Phys.)* **19**, 296 (1930), [arXiv:quant-ph/9903100](https://arxiv.org/abs/quant-ph/9903100).

[7] P. Jordan, J. v. Neumann, and E. Wigner, *Ann. Math.* **35**, 29 (1934).

[8] P. Busch, T. Heinonen, and P. Lahti, *Phys. Rep.* **452**, 155 (2007).

[9] E. Arthurs and J. J. L. Kelly, *Bell Syst. Tech. J.* **44**, 725 (1965).

[10] M. Ozawa, *Phys. Rev. Lett.* **60**, 385 (1988).

- [11] M. Ozawa, [arXiv:1308.3540](#).
- [12] E. Arthurs and M. S. Goodman, *Phys. Rev. Lett.* **60**, 2447 (1988).
- [13] H. Martens and W. M. Muynck, *Found. Phys.* **20**, 255 (1990).
- [14] S. Ishikawa, *Rep. Math. Phys.* **29**, 257 (1991).
- [15] V. Braginski and F. Khalili, *Quantum Measurement* (Cambridge University Press, Cambridge, 1992).
- [16] G. Jaeger, A. Shimony, and L. Vaidman, *Phys. Rev. A* **51**, 54 (1995).
- [17] D. M. Appleby, *Int. J. Theor. Phys.* **37**, 1491 (1998).
- [18] D. M. Appleby, *Int. J. Theor. Phys.* **37**, 2557 (1998).
- [19] W. M. De Muynck, *Found. Phys.* **30**, 205 (2000).
- [20] A. Trifonov, G. Björk, and J. Söderholm, *Phys. Rev. Lett.* **86**, 4423 (2001).
- [21] M. Ozawa, *Phys. Rev. A* **67**, 042105 (2003).
- [22] M. J. W. Hall, *Phys. Rev. A* **69**, 052113 (2004).
- [23] M. Ozawa, *Phys. Lett. A* **320**, 367 (2004).
- [24] Notably, Hall's result uses the standard deviations of optimal estimations rather than the intrinsic preparation deviations σ_A and σ_B , so is in fact a stronger inequality than the one derived by Ozawa.
- [25] H. F. Hofmann, *Phys. Rev. A* **67**, 022106 (2003).
- [26] R. F. Werner, *Quantum Inf. Comput.* **4**, 546 (2004).
- [27] K. Koshino and A. Shimizu, *Phys. Rep.* **412**, 191 (2005).
- [28] P. Busch and D. B. Pearson, *J. Math. Phys.* **48**, 082103 (2007).
- [29] Yu Watanabe, T. Sagawa, and M. Ueda, *Phys. Rev. A* **84**, 042121 (2011).
- [30] Yu Watanabe and M. Ueda, [arXiv:1106.2526](#).
- [31] H. F. Hofmann, [arXiv:1205.0073](#).
- [32] P. Busch, P. Lahti, and R. F. Werner, *Phys. Rev. Lett.* **111**, 160405 (2013).
- [33] P. Busch, P. Lahti, and R. F. Werner, [arXiv:1312.4393](#).
- [34] M. Ozawa, *Ann. Phys.* **311**, 350 (2004).
- [35] J. Erhart, S. Sponar, G. Sulyok, G. Badurek, M. Ozawa, and Y. Hasegawa, *Nat. Phys.* **8**, 185 (2012).
- [36] G. Sulyok, S. Sponar, J. Erhart, G. Badurek, M. Ozawa, and Y. Hasegawa, *Phys. Rev. A* **88**, 022110 (2013).
- [37] So-Young Baek, F. Kaneda, M. Ozawa, and K. Edamatsu, *Sci. Rep.* **3**, 2221 (2013).
- [38] A. P. Lund and H. M. Wiseman, *New J. Phys.* **12**, 093011 (2010).
- [39] Y. P. Terletsky, *Zh. Eksp. Teor. Fiz.* **7**, 1290 (1937).
- [40] H. Margenau and R. N. Hill, *Prog. Theor. Phys.* **26**, 722 (1961).
- [41] L. M. Johansen, *Phys. Lett. A* **329**, 184 (2004).
- [42] Y. Aharonov, D. Z. Albert, and L. Vaidman, *Phys. Rev. Lett.* **60**, 1351 (1988).
- [43] I. M. Duck, P. M. Stevenson, and E. C. G. Sudarshan, *Phys. Rev. D* **40**, 2112 (1989).
- [44] J. Dressel, S. Agarwal, and A. N. Jordan, *Phys. Rev. Lett.* **104**, 240401 (2010).
- [45] L. A. Rozema, A. Darabi, D. H. Mahler, A. Hayat, Y. Soudagar, and A. M. Steinberg, *Phys. Rev. Lett.* **109**, 100404 (2012).
- [46] F. Kaneda, So-Young Baek, M. Ozawa, and K. Edamatsu, [arXiv:1308.5868](#).
- [47] M. M. Weston, M. J. W. Hall, M. S. Palsson, H. M. Wiseman, and G. J. Pryde, *Phys. Rev. Lett.* **110**, 220402 (2013).
- [48] C. Branciard, *Proc. Natl. Acad. Sci. USA* **110**, 6742 (2013).
- [49] C. Branciard, [arXiv:1312.1857](#).
- [50] E. B. Davies and J. T. Lewis, *Commun. Math. Phys.* **17**, 239 (1970).
- [51] M. Ozawa, *J. Math. Phys.* **25**, 79 (1984).
- [52] J. Dressel and A. N. Jordan, *Phys. Rev. A* **85**, 022123 (2012).
- [53] J. Dressel and A. N. Jordan, *Phys. Rev. A* **88**, 022107 (2013).
- [54] J. S. Bell, *Physics* **1**, 195 (1964).
- [55] J. S. Bell, *Rev. Mod. Phys.* **38**, 447 (1966).
- [56] S. Kochen and E. P. Specker, *J. Math. Mech.* **17**, 59 (1967).
- [57] R. W. Spekkens, *Phys. Rev. A* **71**, 052108 (2005).
- [58] R. W. Spekkens, *Phys. Rev. Lett.* **101**, 020401 (2008).
- [59] M. Ozawa, *Phys. Lett. A* **335**, 11 (2005).
- [60] The detector must be initially uncorrelated so it can be independently prepared for each trial and subsequently coupled to different system states.
- [61] M. Ozawa, *J. Opt. B* **7**, S672 (2005).
- [62] A POM also has the common name of positive-operator-valued measure.
- [63] P. A. M. Dirac, *Rev. Mod. Phys.* **17**, 195 (1945).
- [64] S. Chaturvedi, E. Ercolessi, G. Marmo, G. Morandi, N. Mukunda, and R. Simon, *J. Phys. A: Math. Gen.* **39**, 1405 (2006).
- [65] J. S. Lundeen and C. Bamber, *Phys. Rev. Lett.* **108**, 070402 (2012).
- [66] C. Mehta, *J. Math. Phys.* **5**, 677 (1964).
- [67] This expansion generalizes the one used in Ref. [38], which assumes $\hat{M}[m]$ must share the same spectrum with \hat{A} .
- [68] H. M. Wiseman, *Phys. Rev. A* **65**, 032111 (2002).
- [69] J. Dressel and A. N. Jordan, *Phys. Rev. A* **85**, 012107 (2012).
- [70] J. Dressel and A. N. Jordan, *Phys. Rev. Lett.* **109**, 230402 (2012).
- [71] A. G. Kofman, S. Ashhab, and F. Nori, *Phys. Rep.* **520**, 43 (2012).
- [72] J. Dressel, M. Malik, F. M. Miatto, A. N. Jordan, and R. W. Boyd, [arXiv:1305.7154](#).
- [73] G. J. Pryde, J. L. O'Brien, A. G. White, T. C. Ralph, and H. M. Wiseman, *Phys. Rev. Lett.* **94**, 220405 (2005).
- [74] J. Dressel, C. J. Broadbent, J. C. Howell, and A. N. Jordan, *Phys. Rev. Lett.* **106**, 040402 (2011).
- [75] M. E. Goggin, M. P. Almeida, M. Barbieri, B. P. Lanyon, J. L. O'Brien, A. G. White, and G. J. Pryde, *Proc. Natl. Acad. Sci. USA* **108**, 1256 (2011).
- [76] M. Iinuma, Y. Suzuki, G. Taguchi, Y. Kadoya, and H. F. Hofmann, *New J. Phys.* **13**, 033041 (2011).
- [77] J. Dressel, Y. Choi, and A. N. Jordan, *Phys. Rev. B* **85**, 045320 (2012).
- [78] Y. Suzuki, M. Iinuma, and H. F. Hofmann, *New J. Phys.* **14**, 103022 (2012).
- [79] Y. Ota, S. Ashhab, and F. Nori, *Phys. Rev. A* **85**, 043808 (2012).
- [80] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).
- [81] G. Lindblad, *Commun. Math. Phys.* **48**, 119 (1976).
- [82] H. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (Oxford University Press, Oxford, 2007).
- [83] S. M. Barnett, D. T. Pegg, and J. Jeffers, *J. Mod. Opt.* **47**, 1779 (2000).

- [84] D. Pegg, S. Barnett, and J. Jeffers, *J. Mod. Opt.* **49**, 913 (2002).
- [85] T. Amri, J. Laurat, and C. Fabre, *Phys. Rev. Lett.* **106**, 020502 (2011).
- [86] L. A. Rozema, D. H. Mahler, A. Hayat, and A. M. Steinberg, [arXiv:1307.3604](https://arxiv.org/abs/1307.3604).
- [87] A. Danan, D. Farfurnik, S. Bar-Ad, and L. Vaidman, *Phys. Rev. Lett.* **111**, 240402 (2013).