Chapman University
Chapman University Digital Commons

Mathematics, Physics, and Computer Science
Faculty Articles and Research

Science and Technology Faculty Articles and
Research

2-27-2012

Contextual-Value Approach to the Generalized Measurement of Observables

Justin Dressel
Chapman University, dressel@chapman.edu

Andrew N. Jordan
University of Rochester

Follow this and additional works at: http://digitalcommons.chapman.edu/scs_articles

Part of the Quantum Physics Commons

Contextual-Value Approach to the Generalized Measurement of Observables

Contextual-value approach to the generalized measurement of observables

J. Dressel and A. N. Jordan

Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627, USA

We present a detailed motivation for and definition of the *contextual values* of an observable, which were introduced by Dressel *et al*. [Phys. Rev. Lett. **104**, 240401 (2011)]. The theory of contextual values is a principled approach to the generalized measurement of observables. It extends the well-established theory of generalized state measurements by bridging the gap between partial state collapse and the observables that represent physically relevant information about the system. To emphasize the general utility of the concept, we first construct the full theory of contextual values within an operational formulation of classical probability theory, paying special attention to observable construction, detector coupling, generalized measurement, and measurement disturbance. We then extend the results to quantum probability theory built as a superstructure on the classical theory, pointing out both the classical correspondences to and the full quantum generalizations of both Lüder's rule and the Aharonov-Bergmann-Lebowitz rule in the process. As such, our treatment doubles as a self-contained pedagogical introduction to the essential components of the operational formulations for both classical and quantum probability theory. We find in both cases that the contextual values of a system observable form a generalized spectrum that is associated with the independent outcomes of a partially correlated and generally ambiguous detector; the eigenvalues are a special case when the detector is perfectly correlated and unambiguous. To illustrate the approach, we apply the technique to both a classical example of marble color detection and a quantum example of polarization detection. For the quantum example we detail two devices: Fresnel reflection from a glass coverslip, and continuous beam displacement from a calcite crystal. We also analyze the three-box paradox to demonstrate that no negative probabilities are necessary in its analysis. Finally, we provide a derivation of the quantum weak value as a limit point of a pre- and postselected conditioned average and provide sufficient conditions for the derivation to hold.

I. INTRODUCTION

Since the advent of quantum mechanics, practitioners have struggled with an inherent conceptual dualism in its formalism. On one hand, time evolution of a quantum state is a continuous, deterministic, and reversible process well described by a wave equation. On the other hand, there is irreducible stochasticity present in the measurement process that leads to discontinuous and generally irreversible state evolution in the form of so-called "quantum jumps" or "state collapse."

To cope with the necessary introduction of the stochastic element of the theory while still preserving ties with the deterministic classical mechanics, traditional quantum mechanics [1,2] emphasizes the role of Hermitian observable operators that are analogous to classical observables. Indeed, we find that observables underlie most of the core concepts in the quantum theory: commutation relations of observables, complete sets of commuting observables, spectral expansions of observables, conjugate pairs of observables, expectation values of observables, uncertainty relations between observables, and time evolution generated by a Hamiltonian observable. Even the quantum state is introduced as a superposition of observable eigenvectors. The stochasticity of the theory manifests itself as a single prescription for how to average the omnipresent observables under a deterministically evolving quantum state: the implicit projective quantum jumps corresponding to laboratory measurements are largely hidden by the formalism.

Experimental control of quantum systems has improved since the early days of quantum mechanics, however, so the discontinuous evolution present in the measurement process can now be more carefully investigated. Modern optical and condensed matter systems, for example, can condition the evolution of a state on the outcomes of weakly coupled measurement devices (e.g., [3–5]), resulting in *nonprojective* quantum jumps that alter the state more gently, or even resulting in continuous controlled evolution of the state. Since observables are defined in terms of projective jumps that strongly affect the state, it becomes unclear how to correctly apply a formalism based on observables to such nonprojective measurements. A refinement of the traditional formalism must be employed to correctly describe the general case.

To address this need, the theory of *quantum operations*, or generalized measurement, was introduced in the early 1970s by Davies [6] and Kraus [7], and has been developed over the past forty years to become a comprehensive and mathematically rigorous theory [8–16]. The formalism of quantum operations has seen the most use in quantum optics, quantum computation, and quantum information communities, where it is indispensable and well supported by experiment. However, it has not yet seen wide adoption outside of those communities.

Unlike the traditional observable formalism, the formalism of quantum operations emphasizes the *states*. Observables are mentioned infrequently in the quantum operations literature, usually appearing only in the context of projective measurements where they are well understood. Some references (e.g., [13,14,16]) define "generalized observables" in terms of the generalized measurements and detector outcome labels, but give no indication about their relationship to traditional observables, if any. As a result, there is a conceptual gap between the traditional quantum mechanics of observables and

the modern treatment of quantum operations that encompasses a much larger class of possible measurements than the traditional observables seemingly allow.

A possible response to this conceptual gap is to declare that traditional observables are meaningless outside the context of projective measurements. This argument is supported by the fact that any generalized measurement can be understood as a part of a projective measurement being made on a larger joint system that can be associated with a traditional observable in the usual way (i.e., [13, p. 20]). However, there has been parallel research into the "weak measurement" of observables [17–33] that suggests that linking generalized measurements to traditional observables may not be such an outlandish idea.

Weak measurements were introduced as a consequence of the von Neumann measurement protocol [2] that uses an interaction Hamiltonian with variable coupling strength to correlate an observable of interest to the generator of translations for a continuous meter observable. The resulting shift in the meter observable is then used to infer information about the observable of interest in a nonprojective manner. The technique has been used to great effect in the laboratory [34–48] to measure physical quantities like pulse delays, beam deflections, phase shifts, polarization, and averaged trajectories. Therefore, we conclude that there must be some meaningful way to reconcile nonprojective measurements with traditional observables more formally.

The primary purpose of the present work is to detail a synthesis between generalized measurements and observables that is powerful enough to encompass projective measurements, weak measurements, and any strength of measurement in between. The formalism of *contextual values*, which we explicitly introduced in Ref. [49,50] and further developed in [51–53], forms a bridge between the traditional notion of an observable and the modern theory of quantum operations. For a concise introduction to the topic in the context of the quantum theory, we recommend reading our letter [49].

The central idea of the contextual-value formalism is that an observable can be completely measured indirectly using an imperfectly correlated detector by assigning an appropriate set of values to the detector outcomes. The assigned set of values generally differs from the set of eigenvalues for the observable and forms a *generalized spectrum* that is associated with the operations of the generalized measurement, rather than the spectral projections for the observable. Thus, the spectrum that one associates with an observable will depend on the *context* of how the measurement is being performed; such an inability to completely discuss observables without specifying the full measurement context is reminiscent of Bell-Kochen-Specker contextuality [26,54–59] and motivates the name "contextual values."

The secondary purpose of the present work is to demonstrate that the contextual-value formalism for generalized observable measurement is essentially classical in nature. Hence, it has potential applications outside the usual scope of the quantum theory. Indeed, we will show that any system that can be described by Bayesian probability theory can benefit from the contextual-value formalism.

Extending contextual values to the quantum theory from the classical theory clarifies which features of the quantum theory are novel. The quantum theory can be seen as an extension of a classical probability space to a continuous manifold of incompatible frameworks, where each framework is a copy of the original probability space. Hence, intrinsically quantum features arise not from the observables defined in any particular framework, but instead from the relative orientations of the incompatible frameworks. As we shall see, the differences manifest in sequential measurements and conditional measurements due to the probabilistic structure of the incompatible frameworks, rather than the observables or contextual values themselves.

To keep the paper self-contained with these aims in mind, we first develop both the operational approach to measurement and the contextual-value formalism completely within the confines of classical probability theory, giving illustrative examples to cement the ideas. We then port the formalism to the quantum theory and identify the essential differences that arise. Our analysis therefore doubles as a pedagogical introduction to the operational approaches for both classical and quantum probability theory that should be accessible to a wide audience.

The paper is organized as follows: In Sec. I A, we provide a simple intuitive example to introduce the concept of contextual values. In Secs. II A through II C, we develop the classical version of the operational approach to measurement. In Sec. II D, we introduce the contextual-value formalism classically and then give several examples similar to the initial example. In Secs. III A through III C, we generalize the classical operations to quantum operations and highlight the key differences with explicit examples. In Sec. III D, we apply the contextual-value formalism to the quantum case and show that it is unchanged. We also specifically address how to treat weak measurements as a special case of our more general formalism and provide a derivation of the quantum weak value in Sec. III E. Finally, we give our conclusions in Sec. IV.

A. EXAMPLE: COLORBLIND DETECTOR

The idea of the contextual-values formalism is deceptively simple. Its essence can be distilled from the following classical example of an *ambiguous detector*: Suppose we wish to measure a marble that may be colored either red or green. A person with normal vision can distinguish the colors unambiguously and so would represent an ideal detector for the color state of the marble. A partially colorblind person, however, may only estimate the color correctly some percentage of the time and so would represent an ambiguous detector of the color state of the marble.

If the person is only mildly colorblind, then the estimations will be strongly correlated to the actual color of the marble. The ambiguity would then be perturbative and could be interpreted as *noise* introduced into the measurement. However, if the person is strongly colorblind, then the estimations may be only mildly correlated to the actual color of the marble. The ambiguity becomes *nonperturbative*, so the noise dominates the signal in the measurement.

We can design an experimental protocol where an experimenter holds up a marble and the colorblind person gives a thumbs-up if he thinks the marble is green or a thumbs-down if he thinks the marble is red. Suppose, after testing a large number of known marbles, the experimenter determines that

a green marble correlates with a thumbs-up 51% of the time, while a red marble correlates with a thumbs-down 53% of the time. The experimental outcomes of thumbs-up and thumbs-down are thus only weakly correlated with the actual color of the marble.

Having characterized the detector in this manner, the experimenter provides the colorblind person with a very large bag of an unknown distribution of colored marbles. The colorblind person examines every marble and, for each one, records a thumbs-up or a thumbs-down on a sheet of paper, which he then returns to the experimenter. The experimenter then wishes to reconstruct what the average distribution of marble colors in the bag must be, given only the ambiguous output of his colorblind detector.

For simplicity, the clever experimenter decides to associate the colors with numerical values: 1 for green (g) and -1 for red (r). In order to compare the ambiguous outputs with the colors, he also assigns them *different* numerical values: a for thumbs-up (u), and b for thumbs-down (d). He then writes down the following probability constraint equations for obtaining the average marble color, $\langle \text{color} \rangle$, based on what he has observed,

$$
\begin{aligned}
P(u) &= (0.51)P(g) + (0.49)P(r), \\
P(d) &= (0.47)P(g) + (0.53)P(r), \\
\langle \text{color} \rangle &= 1P(g) - 1P(r) = aP(u) + bP(d),
\end{aligned} \tag{1}
$$

which he can rewrite as a matrix equation in the basis of the color probabilities $P(g)$ and $P(r)$,

$$
\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0.51 & 0.47 \\ 0.49 & 0.53 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}. \tag{2}
$$

After solving this equation, he finds that he must assign the amplified values $a = 25$ and $b = -25$ to the outcomes of thumbs-up and thumbs-down, respectively, in order to compensate for the detector ambiguity. After doing so, he can confidently calculate the average color of the marbles in the large unknown bag using the identity (1).

The classical color observable has eigenvalues of 1 and -1 that correspond to an ideal measurement. The amplified values of 25 and -25 that must be assigned to the ambiguous detector outcomes are *contextual values* for the same color observable. The *context* of the measurement is the characterization of the colorblind detector, which accounts for the degree of colorblindness. The expansion (1) relates the spectrum of the observable to its generalized spectrum of contextual values. With this identity, both an ideal detector and a colorblind detector can measure the same observable; however, the assigned values must change depending on the context of the detector being used.

II. CLASSICAL PROBABILITY THEORY

To define contextual values more formally, we shall define generalized measurements within the classical theory of probability using the same language as quantum operations. In particular, rather than representing the observables of classical probability theory in the traditional way as functions, we shall adopt a more calculationally flexible, yet equivalent, *algebraic* representation that closely resembles the operator algebra for quantum observables.

We also briefly comment that the relevant subset of probability theory that is summarized here may slightly differ in emphasis from incarnations that the reader may have encountered previously. Our treatment acknowledges that probability theory, in its most general incarnation, is a system of formal reasoning about Boolean logic propositions [60,61]; specifically, our treatment emphasizes logical inference rather than the traditional frequency analysis of concrete random variable realizations. However, the "frequentist" approach of random variables is not displaced by the logical approach but is rather subsumed as an important special case pertaining to repeatable experiments with logically independent outcomes. Due to its clarity and generality, the logical approach has been widely adopted in diverse disciplines under the distinct name "Bayesian probability theory." Several physicists, including (but certainly not limited to) Jaynes [62], Caves [63], Fuchs [64], Spekkens [65], Harrigan [66], and Leifer [16,67], have also extolled its virtues in recent years. We follow suit to emphasize the generality of the contextual-value concept.

A. Sample spaces and observables

In what follows, we shall consider the stage on which classical probability theory unfolds—namely its space of observables—to be a commutative algebra over the reals that we denote $\Sigma_X^{\mathbb{R}}$. This choice of notation is motivated by the fact that the observable algebra is built from and contains two related spaces, X and Σ_X, that are conceptually distinct and equally important to the theory. The three are illustrated in Fig. 1 to orient the discussion. To avoid distracting technical detail, we will briefly describe finite-dimensional versions of these three spaces here, and note straightforward generalizations to the continuous case when needed [68].

Sample spaces. The core of a probability space is a set of mutually exclusive logic propositions X, known as the *sample space* of *atomic propositions*. In other words, elements of the sample space, such as $g, r \in X$, represent "yes or no" questions that cannot be answered "yes" simultaneously and cannot be broken into simpler questions. For example, $g =$ Does the marble look green? and $r =$ Does the marble look red? are valid mutually exclusive atomic propositions. To be a proper sample space, the propositions should form a complete set, meaning that there must always be exactly one true proposition. Physically, such propositions
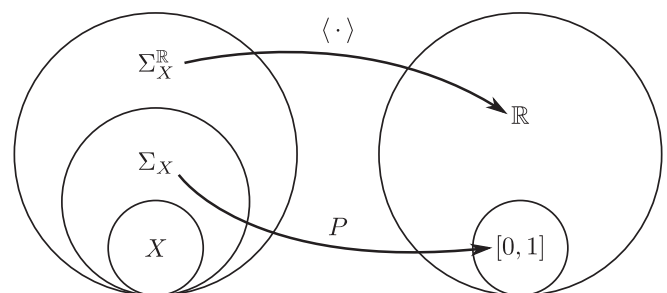
FIG. 1. Diagram of the relationship between the sample space of atomic propositions X, the Boolean algebra of propositions Σ_X, and the algebra of observables $\Sigma_X^{\mathbb{R}}$. The probability state P is a measure from Σ_X to the interval $[0,1]$. The expectation functional $\langle \cdot \rangle$ is a linear extension of P that maps $\Sigma_X^{\mathbb{R}}$ to the reals $\mathbb{R}$; by construction $\langle \cdot \rangle = P(\cdot)$ whenever both are defined.

typically correspond to mutually independent outcomes of an experiment that probes some system of interest. Indeed, any accessible physical property must be testable by some experiment, and any experiment can be described by such a collection of yes or no questions.

Boolean algebra. The atomic propositions in X can be extended to more complex propositions by logical combination in order to form the larger space Σ_X. Specifically, we can combine them algebraically with a logical OR denoted by addition and a logical AND denoted by multiplication. For example, given propositions $x, y, z \in \Sigma_X$, the quantity $xy + yz$ would denote the proposition "(x AND y) OR (y AND z)." Importantly, both the sum and the product commute since the corresponding logical operations commute, and the propositions are idempotent so $x^2 = x$ for any $x \in \Sigma_X$. Furthermore, the product of any two nonequal propositions in X must be trivially false since they are mutually exclusive; we denote the trivially false proposition as 0 since its product with any proposition is also trivially false. Similarly, the sum of all propositions in X will be trivially true since one of the atomic propositions must be true by construction; we denote the trivially true proposition as 1_X since its product with any proposition $x \in X$ leaves that proposition invariant, $1_X x = x$. The logical operation of NOT, or complementation (x^c) with respect to X, can then be defined as the subtraction from the identity $x^c = 1_X - x$ since $x + x^c = 1_X$ must be true for any proposition $x \in X$ by definition. The proposition space Σ_X contains X and is closed under the operations of AND, OR, and NOT; hence, it forms a *Boolean logic algebra* [69].

Observables. Finally, we extend Σ_X linearly over the real numbers to obtain the commutative algebra of *observables* $\Sigma_X^{\mathbb{R}}$. That is, any linear combination of propositions $F = ax + by$ with $a, b \in \mathbb{R}$ and $x, y \in \Sigma_X$ is an observable in $\Sigma_X^{\mathbb{R}}$; similarly any linear combination of observables $H = a'F + b'G$ with $a', b' \in \mathbb{R}$ and $F, G \in \Sigma_X^{\mathbb{R}}$ is also an observable in $\Sigma_X^{\mathbb{R}}$. Countable sums are permitted provided the coefficients converge. The three spaces X, Σ_X, and $\Sigma_X^{\mathbb{R}}$ are illustrated in Fig. 1.

The observables combine logical propositions with numbers that describe the relation of each proposition to some meaningful reference. For example, one could define a simple observable $A = (1)g + (-1)r$ that assigns a value of 1 to the proposition asking whether a marble looks green and assigns a value of -1 to the proposition asking whether that same marble looks red in order to distinguish the colors by a sign. Alternatively, one can bestow a physical meaning to the color propositions by defining a wavelength observable instead: $B = (550 \text{ nm})g + (700 \text{ nm})r$. One could even define an observable $C = (\$2)g + (-\$3)r$ that indicates a monetary bet made on the color of the marble, with \$2 being awarded for a color of green and \$3 being lost for a color of red. Such numerical labels are always assigned by convention, but indicate physically relevant information about the type of questions being asked by the experimenter that are answerable by the independent propositions.

Representation. The algebra Σ_X can be represented as the lattice of projection operators acting on a Hilbert space exactly as in the standard representation of quantum theory [2,12,70]. The elements $\{x\}$ of X correspond to rank-1 projection operators $\{|x\rangle\langle x|\}$ onto orthogonal subspaces spanned by orthonormal vectors $\{|x\rangle\}$ in the Hilbert space. Any sum

of n elements of X, $x_1 + \cdots + x_n$, corresponds to a rank-n projection operator $|x_1, \ldots, x_n\rangle\langle x_1, \ldots, x_n|$ onto a subspace spanned by n orthonormal vectors $\{|x_1\rangle, \ldots, |x_n\rangle\}$ in the Hilbert space. Hence, we shall casually refer to propositions of the Boolean algebra Σ_X as *projections* in what follows. However, it is important to note that the Boolean algebra Σ_X need not be represented in this fashion to be well defined.

Just like the propositions Σ_X can be represented as projections on a Hilbert space, the observables $\Sigma_X^{\mathbb{R}}$ can also be represented as the algebra of Hermitian operators acting on the same Hilbert space. Hence, we shall casually refer to observables in $\Sigma_X^{\mathbb{R}}$ as *observable operators* in analogy to the quantum theory. However, unlike quantum observables, all classical observables commute. It is important to note that the representation of observables as operators on a Hilbert space in both the classical and the quantum case remains strictly optional for calculational convenience.

Independent probability observables. We note that the identity observable 1_X can be partitioned into many distinct sets of independent propositions in Σ_X, such as $\sum_i x_i = 1_X$, which is known as a *closure relation*. Each partitioning corresponds to a particular detector arrangement that only probes those propositions. Such a partitioning $\{x_i\}$ has the common mathematical name *projection-valued measure* (PVM) since it forms a measure over the index i and has a representation that consists of orthogonal projections. However, we shall make an effort to call the propositions $\{x_i\}$ *independent probability observables* to be more physically descriptive. We will later contrast them with more general probability observables.

General observables can be constructed from independent probability observables by associating a real value $f(x_i)$ to each index i in the sum, $F = \sum_i f(x_i)x_i$. The product of the observable with any of its constituent probability observables simplifies, $Fx_i = f(x_i)x_i$; hence, the associated values form the set of *eigenvalues* for the observable. For a finite observable space $\Sigma_X^{\mathbb{R}}$, the set of atomic propositions X itself is a maximally refined set of independent probability observables that can construct any observable in the space,

$$F = \sum_{x \in X} f(x)x. \tag{3}$$

In the continuous case the values $f(x)$ form a measurable function that specifies the *spectrum* of the observable; the sum (3) is then commonly written as an integral over the continuous set of propositions $\{|x\rangle\langle x|\}$, $F = \int_X f(x)d|x\rangle\langle x|$. We use the Hilbert space notation $d|x\rangle\langle x|$ in the integral to avoid later confusion with real-valued integrals.

B. States, densities, and collapse

Probability measures. A *state* P is a *probability measure* over the Boolean algebra Σ_X, meaning that it is a linear map from Σ_X to the interval $[0,1]$ such that $P(1_X) = 1$. Such a state P assigns a numerical value $P(x)$ to each proposition $x \in \Sigma_X$ that quantifies its degree of *plausibility*; that is, $P(x)$ formally indicates how likely it is that the question x would be answered "yes" were it to be answered, with 1 indicating a certain "yes" and 0 indicating a certain "no." The value $P(x)$ is called the *probability* for the proposition x to be true. Normalizing $P(1_X) = 1$ ensures that exactly one proposition

in the sample space must be true. For continuous spaces, the state becomes an integral $P(x_0) = \int_{x_0 \in \Sigma_X} dP(x)$.

Frequencies. Empirically, one can check probabilities by repeatedly asking a proposition in Σ_X to identically prepared systems and collecting statistics regarding the answers. For a particular proposition $x \in \Sigma_X$, the ratio of yes answers to the number of trials will converge to the probability $P(x)$ as the number of trials becomes infinite. However, the probability has a well-defined meaning as a plausibility prediction even without actually performing such a repeatable experiment. Indeed, designing good quality repeatable experiments to check the probabilities assigned by a predictive state is the primary goal of experimental science, and is generally quite difficult to achieve.

Expectation functionals. The linear extension of a state P to the whole observable algebra $\Sigma_X^{\mathbb{R}}$ is an *expectation functional* that averages the observables and is traditionally notated with angled brackets $\langle \cdot \rangle$. Specifically, for an observable $F = \sum_{x \in X} f(x)x$, then,

$$\langle F \rangle = \sum_{x \in X} f(x)P(x), \tag{4}$$

is the *expectation value*, or average value, of F under the functional $\langle \cdot \rangle$ that extends the probability state P. Since $\langle \cdot \rangle$ is linear, it passes through the sum and the constant factors of $f(x)$ to apply directly to the propositions x. The restriction of $\langle \cdot \rangle$ to Σ_X is P, so $\langle x \rangle = P(x)$ as written in Eq. (4). That is, the expectation value $\langle x \rangle$ of a pure proposition x is the probability of that proposition. The probability state P and its linear extension $\langle \cdot \rangle$ are illustrated in Fig. 1. For continuous spaces the sum (4) becomes an integral of the measurable function $f(x)$, $\langle F \rangle = \langle \int_X f(x)d|x\rangle\langle x| \rangle = \int_X f(x)dP(x)$.

Moments. The nth statistical moment of F is $\langle F^n \rangle = \sum_{x \in X} f^n(x)P(x)$ and empirically corresponds to measuring the observable F n times in a row per trial on identical systems and averaging the repeated results. Hence, the moments quantify the fluctuations of the observable measurements that stem from uncertainty in the state. For continuous spaces, the higher moments also become integrals $\langle F^n \rangle = \int_X f^n(x)dP(x)$.

Densities. States can often be represented as *densities* with respect to some *reference measure* μ from Σ_X to $\mathbb{R}^+$, which can be convenient for calculational purposes. Just as the state P can be linearly extended to an expectation functional $\langle \cdot \rangle$, any reference measure μ can be linearly extended to a functional $\langle \cdot \rangle_\mu$. For continuous spaces, such a reference functional takes the form of an integral $\langle F \rangle_\mu = \int_X f(x)d\mu(x)$. The representation of a state as a density follows from changing the integration measure for the state to the reference measure $\langle F \rangle = \int_X f(x)dP(x) = \int_X f(x)(dP/d\mu)(x)d\mu(x)$. The Jacobian conversion factor $dP/d\mu$ from the integral over $dP(x)$ to the integral over a different measure $d\mu(x)$ is the *probability density* for P with respect to μ, if it exists [71]. We can then define a *state density observable* $P_\mu = \int_X (dP/d\mu)(x)d|x\rangle\langle x|$ that relates the expectation functional $\langle \cdot \rangle$ to the reference functional $\langle \cdot \rangle_\mu$ directly according to the relation $\langle P_\mu F \rangle_\mu = \langle F \rangle$.

For continuous spaces, the standard integral is most frequently used as a reference. Hence, the probability density with respect to the standard integral is given the simple notation $p(x)$ such that $\langle F \rangle = \int_X f(x)p(x)dx$. Importantly,

the probability for x is not the density $p(x) = (dP/dx)(x)$ but is the (generally infinitesimal) integral of the density over a single point $P(x) = \int_{x \in X} p(x)dx$ [72,73], commonly notated $p(x)dx$.

In discrete spaces we apply the same idea by defining a state-density observable directly in terms of measure ratios,

$$P_\mu = \sum_{x \in X} \frac{P(x)}{\mu(x)} x. \tag{5}$$

Then by definition and linearity, $\langle P_\mu F \rangle_\mu = \sum_{x \in X}[P(x)/\mu(x)]f(x)\mu(x) = \sum_{x \in X} f(x)P(x) = \langle F \rangle$, as required. Evidently, the measure μ must be nonzero for all propositions x for which P is nonzero in order for such a state density to be well defined. This definition as a ratio of functionals will correctly reproduce the Jacobian derivative in the continuous case using a limiting prescription.

Trace. An important reference measure which is nonzero for any nonzero proposition is the counting measure, or *trace* Tr, which evaluates to the rank of any proposition in Σ_X; for example, given $x, y, z \in X$ then $(x + y + z) \in \Sigma_X$ is a rank-3 proposition and $\text{Tr}(x + y + z) = \text{Tr}(x) + \text{Tr}(y) + \text{Tr}(z) = 1 + 1 + 1 = 3$. Since the trace evaluates to unity on any atomic proposition, *any* state has a trace density defined by equation (5) that is traditionally notated as ρ:

$$\rho = \sum_{x \in X} P(x)x. \tag{6}$$

The trace density is the only state density that is always defined and exactly determined by the probabilities of the atomic propositions $P(x)$. Because of this, the trace representation of a state can be naturally interpreted as an inner product,

$$\langle \rho, F \rangle = \text{Tr}(\rho F) = \langle F \rangle, \tag{7}$$

between the trace-density and the observable, known as the Hilbert-Schmidt inner product. The trace will become particularly important when we generalize to quantum mechanics, which is why we mention it here. Indeed, the trace-density ρ will be equivalent to the quantum mechanical density operator when extended to the noncommutative case. For continuous spaces the integral is traditionally preferred to the trace as a reference because the trace can frequently diverge.

State collapse. If a question on the probability space is answered by some experiment, then the state indicating the plausibilities for future answers must be updated to reflect the acquired answer. The update process is known as *Bayesian state conditioning*, or *state collapse*. Specifically, if a proposition $y \in \Sigma_X$ is verified to be true, then the experimenter updates the expectation functional to the conditioned functional,

$$\langle F \rangle_y = \frac{\langle yF \rangle}{P(y)}, \tag{8}$$

that reflects the new information. For a proposition $x \in \Sigma_X$, the conditional probability $\langle x \rangle_y = P(yx)/P(y)$ has the traditional notation $P(x|y)$ and is read as "the probability of x given y."

From (8), any state density corresponding to P will be similarly updated to a new density via a product,

$$P_{\mu|y} = \frac{P_\mu y}{P(y)}. \tag{9}$$

Notably, conditioning the trace-density ρ on an atomic proposition $y \in X$ will collapse the density to become the proposition itself, $\rho_y = \rho y / P(y) = y$.

Note that the proposition y serves a dual role in the conditioning procedure. First, it is used to compute the normalization probability $P(y)$. Second, it directly updates the state via a product action. The product indicates that future questions will be logically linked to the answered question with the AND operation; that is, the knowledge about the system has been refined by the answered question. The process of answering a question about the system and then conditioning the state on the new information is called a *measurement*; moreover, since the proposition y is a projection acting on the density, this kind of measurement is called a *projective measurement*.

Bayes' rule. If we pick another proposition $z \in \Sigma_X$ as the observable in Eq. (8) we can derive *Bayes' rule* as a necessary consequence by interchanging y and z and then equating the joint probabilities $P(yz)$,

$$P(z|y) = P(y|z) \frac{P(z)}{P(y)}. \tag{10}$$

Bayes' rule relates conditioned expectation functionals to one another and so is a powerful logical inference tool that drives much of the modern emphasis on the logical approach to probability theory.

Disturbance. Conditioning, however, is not the only way that one can alter a state. One can also *disturb* a state without learning any information about it, which creates a transition to an updated expectation functional that we denote with a tilde $\langle \widetilde{\cdot} \rangle$ according to

$$\langle \widetilde{F} \rangle = \langle \mathcal{D}(F) \rangle, \tag{11a}$$

$$\mathcal{D}(F) = \sum_{x \in X} \langle F \rangle_{D_x} x, \tag{11b}$$

$$\langle F \rangle_{D_x} = \sum_{x' \in X} f(x') D_x(x'). \tag{11c}$$

Here the disturbance $\mathcal{D}$ is a map from $\Sigma_X^{\mathbb{R}}$ to $\Sigma_X^{\mathbb{R}}$ that is governed by a collection of states $\{D_x\}$ that specify *transition probabilities* $D_x(x')$ from old propositions x to new propositions x'. To be normalized, the transition states must satisfy $D_x(1_X) = 1$, so that $\langle 1_X \rangle_{D_x} = 1_X$ and therefore $\mathcal{D}(1_X) = 1_X$. Updating the state according to (11) is also known as *Bayesian belief propagation* [67] and is more commonly written in the fully expanded form $\langle \widetilde{F} \rangle = \sum_{x \in X} P(x) \sum_{x' \in X} D_x(x') f(x')$.

Time evolution. As an important special case, the time evolution of a Markovian stochastic process is a form of disturbance $\mathcal{D}_t$, known as a *propagator*, that is parametrized by a time interval t. No information is learned as the system evolves, so the knowledge about the system as represented by the expectation functional can only propagate according to the laws governing the time evolution. For a Hamiltonian system, the time evolution is of Liouville form; that is, if we define a time-evolving observable as $F(t) = \mathcal{D}_t(F)$ then we have $dF(t)/dt = \{F(t), H\}_p$, where $\{\cdot, \cdot\}_p$ is defined point wise as the Poisson bracket. The differential equation implicitly specifies the form of the disturbance $\mathcal{D}_t$.

Correlation functions. Correlations between observables at different times can be obtained by inserting a time-evolution disturbance between the observable measurements,

$$\langle F(0)G(t) \rangle = \langle F \mathcal{D}_t(G) \rangle$$
$$= \sum_{x \in X} P(x) f(x) \sum_{x' \in X} D_{x,t}(x') g(x'). \tag{12}$$

Operationally, this corresponds to measuring the observable F, waiting an interval of time t, then measuring the observable G. Similarly, n-time correlations can be defined with $n - 1$ time-evolution disturbances between the observable measurements $\langle F_1 \mathcal{D}_{t_1}(F_2 \cdots \mathcal{D}_{t_{n-1}}(F_n) \cdots) \rangle$. Computing the correlation of an observable with itself at the same time will produce a higher moment $\langle F^n \rangle$.

Invasive measurement. A system may also be disturbed during the physical process that implements conditioning, which will alter the state above and beyond the pure conditioning expression (8). With such an *invasive measurement*, one conditions a state after a disturbance induced by the measurement process has occurred; hence, one obtains a new state,

$$\langle \widetilde{F} \rangle_y = \frac{\langle \mathcal{D}(yF) \rangle}{\langle \mathcal{D}(y) \rangle} = \frac{\sum_{x \in X} P(x) \sum_{x' \in X} D_x(yx') f(x')}{\sum_{x \in X} P(x) D_x(y)}, \tag{13}$$

which is a composition of the measurement disturbance (11) followed by the pure conditioning (8). As we shall see later in Sec. III B, the quantum projection postulate (Lüder's Rule) can be understood as an invasive measurement similar to (13), but not as pure conditioning similar to (8). This observation has also been recently emphasized by Leifer and Spekkens [67], who show that a careful extension of (8) to the noncommutative quantum setting does not reproduce the projection postulate. Hence, better understanding classical invasive measurement should provide considerable insight into the quantum measurement process. However, to properly understand the implications of invasive measurements on the measurement of observables, we must consider the measurement process in more detail.

C. Detectors and probability observables

For a single ideal experiment that answers questions of interest with perfectly correlated independent outcomes, knowing the spectrum of an observable for that experiment is completely sufficient. However, in many (if not most) cases the independent propositions corresponding to the experimental outcomes are only *imperfectly correlated* with the questions of interest about the system. Since in such a case one may not have direct access to the questions of interest, one also may not have direct access to the observables of interest. One must instead *infer* information about the observables of interest *indirectly* from the correlated outcomes of the detector to which one does have access.

Joint sample space. To handle this case formally, we first enlarge the sample space to include both the sample space of interest, which we call the *system* X and the accessible sample space, which we call the *detector* Y. Questions about the system and the detector can be asked independently, so every question for the system can be paired with any question from the detector; therefore, the resulting joint sample space

must be a product space, $XY = \{xy | x \in X, y \in Y\}$, where the products of propositions from different sample spaces commute. The Boolean algebra Σ_{XY} and observable algebra $\Sigma_{XY}^{\mathbb{R}}$ are constructed in the usual way from the joint sample space, and contain the algebras Σ_X, Σ_Y, $\Sigma_X^{\mathbb{R}}$, and $\Sigma_Y^{\mathbb{R}}$ as subalgebras. When represented as operators on a Hilbert space, the corresponding joint representation exists within the tensor product of the system and detector space representations.

Product states. If the probabilities of the system propositions are uncorrelated with the probabilities of the detector propositions under a joint state P on the joint sample space, then the joint state can be written as a *composition* of independent states that are restricted to the sample spaces of the system and detector: $P = P_X \circ P_Y$. Just as the state P has a linear extension to $\langle \cdot \rangle$, its restrictions P_X and P_Y have linear extensions $\langle \cdot \rangle_X$ and $\langle \cdot \rangle_Y$, respectively. Thus, for any joint observable F an uncorrelated expectation has the form $\langle F \rangle = \langle \langle F \rangle_Y \rangle_X = \langle \langle F \rangle_X \rangle_Y$. Such an uncorrelated joint state is known as a *product state*. The name stems from the fact that, for a simple product $F_X F_Y$ of system and detector observables, the corresponding joint expectation decouples into a product of system and detector expectations separately, $\langle F_X F_Y \rangle = \langle F_X \rangle_X \langle F_Y \rangle_Y$.

Similarly, general measures on the joint sample space can be product measures. A particularly useful example is the trace $\mathrm{Tr} = \mathrm{Tr}_X \circ \mathrm{Tr}_Y$ on XY, which is composed of the *partial traces* Tr_X and Tr_Y. The trace serves as a convenient reference measure since it is a product measure for which any joint state has a corresponding density. On continuous spaces the standard integral is also a product measure, $\langle F \rangle = \int_X [\int_Y f(x,y) p(x,y) dy] dx = \int_Y [\int_X f(x,y) p(x,y) dx] dy$, which tends to have noninfinitesimal densities.

Correlated states. In addition to product states, the joint space admits a much larger class of *correlated* states where the detector and system questions are dependent on one another. With such a correlated state a measurement on the detector cannot be decoupled in general from a measurement on the system. Information gathered from a measurement on a detector under a correlated state will also indirectly provide information about the system, thus motivating the term "detector."

Reduced states. For a pure system observable F_X or a pure detector observable F_Y, the average under a joint state will be equivalent to the average under a state restricted to either the system or the detector space, known as a *reduced state* or a *marginalized state*. We can define such a reduced state by using the joint state density under any reference *product* measure $\mu = \mu_X \circ \mu_Y$, such as the trace Tr. It then follows that

$$\langle F_X \rangle = \left(\langle P_\mu \rangle_{\mu_Y} F_X \right)_{\mu_X} = \left(P_{\mu_X} F_X \right)_{\mu_X}, \quad (14a)$$

$$\langle F_Y \rangle = \left(\langle P_\mu \rangle_{\mu_X} F_Y \right)_{\mu_Y} = \left(P_{\mu_Y} F_Y \right)_{\mu_Y}. \quad (14b)$$

The quantities $P_{\mu_X} = \langle P_\mu \rangle_{\mu_Y}$ and $P_{\mu_Y} = \langle P_\mu \rangle_{\mu_X}$ are the *reduced state densities* that define the reduced states P_X and P_Y with expectation functionals

$$\langle F_X \rangle_X = \left(P_{\mu_X} F_X \right)_{\mu_X}, \quad (15a)$$

$$\langle F_Y \rangle_Y = \left(P_{\mu_Y} F_Y \right)_{\mu_Y}. \quad (15b)$$

By definition, $\langle F_X \rangle = \langle F_X \rangle_X$ and $\langle F_Y \rangle = \langle F_Y \rangle_Y$. However, in general $\langle F \rangle \neq \langle \langle F \rangle_Y \rangle_X$, $\langle F \rangle \neq \langle \langle F \rangle_X \rangle_Y$, and $\langle \langle F \rangle_Y \rangle_X \neq \langle \langle F \rangle_X \rangle_Y$ unless P is a product state. The resulting reduced expectations $\langle \cdot \rangle_X$ and $\langle \cdot \rangle_Y$ are independent of the choice of reference product functional μ.

Probability observables. Any correlation between the system and detector in the joint state allows us to directly relate propositions on the detector to *observables* on the system. We can compute the relationship directly by using a closure relation and rearranging the conditioning procedure (8) to find

$$P(y) = \sum_{x \in X} P(x) P(y|x) = \left\langle \sum_{x \in X} P(y|x) x \right\rangle = \langle E_y \rangle_X, \quad (16)$$

$$E_y = \sum_{x \in X} P(y|x) x. \quad (17)$$

The resulting set of system observables $\{E_y\}$ exactly correspond to the detector outcomes $\{y\}$. Analogously to a set of independent probability observables, they form a partition of the system identity, but are indexed by detector propositions rather than by system propositions, $\sum_{y \in Y} E_y = 1_X$. Such a set $\{E_y\}$ has the common mathematical name *positive operator-valued measure* (POVM) [11], since it forms a measure over the detector sample space Y consisting of positive operators. However, we shall make an effort to refer to them as general *probability observables* to emphasize their physical significance. As long as the detector outcomes are not mutually exclusive with the system, the probability observables (17) will be a faithful representation of the reduced state of the detector in the observable space of the system.

Process tomography. The probability observables are completely specified by the *conditional likelihoods* $P(y|x)$ for a detector proposition y to be true given that a system proposition x is true. Such conditional likelihoods are more commonly known as *response functions* for the detector and can be determined via independent *detector characterization* using known reduced system states; such characterization is also known as *detector tomography* or *process tomography*. Any good detector will then maintain its characterization with any *unknown* reduced system state. That is, a noninvasive coupling of such a good detector to an unknown system produces a correlated joint state according to $P(xy) = P_X(x) P(y|x)$, where P_X is the unknown reduced system state prior to the interaction with the detector.

Generalized state collapse. In addition to allowing the computation of detector probabilities, $P(y) = \langle E_y \rangle_X$, probability observables also have the dual role of updating the reduced system state following a measurement on the detector. To see this, we apply the general rule for state collapse (8) for a detector proposition y on the joint state to find

$$\langle F_X \rangle_y = \frac{\langle y F_X \rangle}{P(y)} = \sum_{x \in X} f_X(x) P(y|x) \frac{P_X(x)}{P(y)} = \frac{\langle E_y F_X \rangle_X}{\langle E_y \rangle_X}, \quad (18)$$

which can be seen as a generalization of the Bayesian conditioning rule (8) to account for the effect of an imperfectly correlated detector, and it can also be understood as a form of *Jeffrey's conditioning* [74]. For this reason, probability observables are commonly called *effects* of the *generalized*

measurement. A reduced state density P_{μ_X} for the system updates as $P_{\mu_X|y} = P_{\mu_X} E_y / \langle E_y \rangle_X$. Such a *generalized measurement* is nonprojective, so is not constrained to the disjoint questions on the sample space of the system. As a result, it answers questions on the system space *ambiguously* or *noisily*.

Weak measurement. The extreme case of such an ambiguous measurement is a *weak measurement*, which is a measurement that does not (appreciably) collapse the system state. Such a measurement is inherently ambiguous to the extent that only a minuscule amount of information is learned about the system with each detection. Formally, the probability observables for a weak measurement are all nearly proportional to the identity on the system space. Typically, an experimenter has access to some control parameter ϵ (such as the correlation strength) that can alter the weakness of the measurement such that

$$\forall \, y, \lim_{\epsilon \to 0} E_y(\epsilon) = P_Y(y) 1_X, \tag{19}$$

where $P_Y(y) \in (0,1)$ is the nonzero probability of obtaining the detector outcome y in the absence of any interaction with the system. Then, for small values of ϵ, the measurement leaves the system state nearly unperturbed, $P_{\mu_X|y} = P_{\mu_X} E_y(\epsilon) / \langle E_y(\epsilon) \rangle_X \approx P_{\mu_X}$. The limit as such a control parameter $\epsilon \to 0$ is known as the *weak-measurement limit* and is a formal idealization not strictly achievable in an experiment.

Strong measurement. The opposite extreme case is a *strong measurement* or projective measurement, which is a measurement for which all outcomes are independent, as in Eq. (3). In other words, the probability observables are independent for a strong measurement. The projective collapse rule (8) can therefore be seen as a special case of the general collapse rule (18) from this point of view.

Measurement sequences. A further benefit of the probability observable representation of a detector is that it becomes straightforward to discuss sequences of generalized measurements performed on the same system. For example, consider two detectors that successively couple to a system and have the outcomes y and z measured, respectively. To describe the full joint state of the system and both detectors requires a considerably enlarged sample space. However, if the detectors are characterized by two sets of probability observables $\{E_y\}$ and $\{E'_z\}$ we can immediately write down the probability of both outcomes to occur as well as the resulting final collapsed system state without using the enlarged sample space:

$$P(yz) = \langle E'_z E_y \rangle_X, \tag{20a}$$

$$\langle F_X \rangle_{yz} = \frac{\langle E'_z E_y F_X \rangle_X}{\langle E'_z E_y \rangle_X}. \tag{20b}$$

Similarly, a conditioned density takes the form $P_{\mu_X|yz} = P_{\mu_X} E'_z E_y / \langle E'_z E_y \rangle_X$. The detectors have been *abstracted* away to leave only their effect upon the system of interest.

Generalized invasive measurement. The preceding discussion holds provided that the detector can be noninvasively coupled to a reduced system state P_X to produce a joint state $P(xy) = P_X(x) P(y|x)$. However, more generally the process of coupling a reduced detector state P_Y to the reduced system state P_X will *disturb* both states, as discussed for (11). The disturbance produces a joint state from the original product state of the system and detector according to

$$\langle \widetilde{xy} \rangle = \langle \langle \mathcal{D}(xy) \rangle_Y \rangle_X, \tag{21}$$

$$\mathcal{D}(xy) = \sum_{x' \in X} \sum_{y' \in Y} D_{x',y'}(xy) x' y', \tag{22}$$

where $D_{x',y'}$ are states specifying the joint transition probabilities for the disturbance. The noninvasive coupling $P(xy) = P_X(x) P(y|x)$ is a special case of this where the reduced system state is unchanged by the coupling.

As a result, we must slightly modify the derivation of the probability observables (16) to properly include the disturbance,

$$\langle \widetilde{y} \rangle = \langle \langle \mathcal{D}(y) \rangle_Y \rangle_X = \langle \tilde{E}_y \rangle_X, \tag{23a}$$

$$\tilde{E}_y = \langle \mathcal{D}(y) \rangle_Y = \sum_{x \in X} \sum_{y' \in Y} P_Y(y') D_{x,y'}(y) x. \tag{23b}$$

The modified probability observable $\tilde{E}_y$ includes both the initial detector state P_Y and the disturbance from the measurement. Detector tomography will therefore find the effective characterization probabilities $\tilde{P}(y|x) = \sum_{y' \in Y} D_{x,y'}(y) P_Y(y')$.

The generalized collapse rule similarly must be modified to include the disturbance,

$$\langle \widetilde{F_X} \rangle_y = \frac{\langle \langle \mathcal{D}(y F_X) \rangle_Y \rangle_X}{\langle \langle \mathcal{D}(y) \rangle_Y \rangle_X} = \frac{\langle \mathcal{E}_y(F_X) \rangle_X}{\langle \tilde{E}_y \rangle_X}, \tag{24}$$

$$\begin{aligned}
\mathcal{E}_y(F_X) &= \langle \mathcal{D}(y F_X) \rangle_Y \\
&= \sum_{x' \in X} x' \sum_{y' \in Y} P_Y(y') \sum_{x \in X} D_{x',y'}(yx) f(x). \tag{25}
\end{aligned}$$

Surprisingly, we can no longer write the conditioning in terms of just the probability observables $\tilde{E}_y$; instead we must use an *operation* $\mathcal{E}_y$ that takes into account both the coupling of the detector and the disturbance of the measurement in an active way. The measurement operation is related to the effective probability observable according to, $\mathcal{E}_y(1_X) = \tilde{E}_y$.

The change from observables to operations when the disturbance is included becomes particularly important for a sequence of invasive measurements. Consider an initial system state P_X that is first coupled to a detector state P_Y via a disturbance $\mathcal{D}_1$, then conditioned on the detector proposition y, then coupled to a second detector state P_Z via a disturbance $\mathcal{D}_2$, and finally conditioned on the detector proposition z. The joint probability for obtaining the ordered sequence (y,z) can be written as

$$\langle \langle \mathcal{D}_1(y \langle \mathcal{D}_2(z) \rangle_Z) \rangle_Y \rangle_X = \langle \mathcal{E}_y(\tilde{E}'_z) \rangle_X. \tag{26}$$

The effective probability observable $\mathcal{E}_y(\mathcal{E}'_z(1_X)) = \mathcal{E}_y(\tilde{E}'_z)$ for the ordered measurement sequence (y,z) is no longer a simple product of the probability observables $\tilde{E}_y$ and $\tilde{E}'_z$ as in Eq. (20a), but is instead an ordered *composition of operations*.

The ordering of operations also leads to a new form of *postselected* conditioning. Specifically, if we condition only

on the second measurement of z in an invasive sequence (y,z), we obtain

$$_z\langle\widetilde{y}\rangle = \frac{\langle\mathcal{E}_y(\tilde{E}'_z)\rangle_X}{\sum_{y'\in Y}\langle\mathcal{E}_{y'}(\tilde{E}'_z)\rangle_X} = \frac{\langle\mathcal{E}_y(\tilde{E}'_z)\rangle_X}{\langle\mathcal{E}(\tilde{E}'_z)\rangle_X}, \qquad (27)$$

$$\mathcal{E}(\tilde{E}'_z) = \sum_{y'\in Y}\mathcal{E}_{y'}(\tilde{E}'_z) = \langle\mathcal{D}(\tilde{E}'_z)\rangle_Y. \qquad (28)$$

The different position of the subscript serves to distinguish the postselected probability $_z\langle\widetilde{y}\rangle$ from the preselected probability $\langle\widetilde{y}\rangle_z = \langle\mathcal{E}'_z(\tilde{E}_y)\rangle_X/\langle\tilde{E}'_z\rangle_X$ corresponding to the reverse measurement ordering of (z,y). The operation $\mathcal{E}$ appearing in the denominator is called a *nonselective measurement* since it includes the disturbance induced by the measurement coupling, but does not condition on any particular detector outcome. When the disturbance to the reduced system state vanishes, the conditioning becomes order independent and both types of conditional probability reduce to $P(y|z) = \langle E_y E'_z\rangle_X/\langle E'_z\rangle_X$.

The two forms of conditioning for invasive measurements in turn lead to a modified form of Bayes' rule that relates the preselected conditioning of a sequence to the postselected conditioning of the same sequence,

$$_z\langle\widetilde{y}\rangle = \langle\widetilde{z}\rangle_y\frac{\langle E_y\rangle_X}{\langle\mathcal{E}(\tilde{E}'_z)\rangle_X}. \qquad (29)$$

When the disturbance to the reduced system state vanishes, the nonselective measurement $\mathcal{E}$ reduces to the identity operation, $_z\langle\widetilde{y}\rangle$ reduces to $P(y|z)$, $\langle\widetilde{z}\rangle_y$ reduces to $P(z|y)$, and we correctly recover the noninvasive Bayes' rule (10).

D. Contextual values

Observable correspondence. With the preliminaries about generalized state conditioning out of the way, we are now in a position to discuss the measurement of observables in more detail. First we observe an important corollary of the observable representation of the detector probabilities $P(y) = \langle E_y\rangle_X$ from (16): *detector* observables can be mapped into equivalent *system* observables,

$$\langle F_Y\rangle = \sum_{y\in Y} f_Y(y)P(y) = \langle F_X\rangle_X, \qquad (30)$$

$$F_X = \sum_{y\in Y} f_Y(y)E_y. \qquad (31)$$

Note that the eigenvalues $f_X(x) = \sum_{y\in Y} f_Y(y)P(y|x)$ of the equivalent system observable F_X are not the same as the eigenvalues $f_Y(y)$ of the original detector observable F_Y, but are instead their average under the detector response. If the system propositions were accessible then the system observable F_X would allow nontrivial inference about the detector observable F_Y, provided that the probability observables were nonzero for all y in the support of F_Y.

Contextual values. A more useful corollary of the expansion (31) is that any *system* observable that can be expressed as a combination of probability observables may be equivalently expressed as a *detector* observable,

$$F_X = \sum_{y\in Y} f_Y(y)E_y \Rightarrow F_Y = \sum_{y\in Y} f_Y(y)y, \qquad (32)$$

which is the classical form of our main result. Using this equivalence, *we can indirectly measure such system observables using only the detector.* We dub the eigenvalues of the detector observable $f_Y(y)$ the contextual values (CVs) of the system observable F_X under the *context of the specific detector* characterized by a specific set of probability observables $\{E_y\}$. The CVs form a *generalized spectrum* for the observable since they are associated with general probability observables for a generalized measurement and not independent probability observables for a projective measurement; the eigenvalues are a special case when the probability observables are the spectral projections of the observable being measured.

With this point of view, we can understand an observable as an *equivalence class* of possible measurement strategies for the same average information. That is, using appropriate pairings of probability observables and CVs, one can measure the same observable average in many different ways, $\langle F_X\rangle = \sum_{x\in X} f_X(x)P(x) = \sum_{y\in Y} f_Y(y)\langle E_y\rangle_X$. Each such expansion corresponds to a different experimental setup.

Moments. Similarly, the nth statistical moment of an observable can be measured in many different, yet equivalent, ways. For instance, the nth moment of an observable F_X can be found from the expansion (32) as

$$\langle(F_X)^n\rangle = \left\langle\left(\sum_{y\in Y} f_Y(y)E_y\right)^n\right\rangle_X$$

$$= \sum_{y_1,\ldots,y_n\in Y} f_Y(y_1)\cdots f_Y(y_n)\langle E_{y_1}\cdots E_{y_n}\rangle_X. \qquad (33)$$

By examining the general collapse rule for measurement sequences (20a) we observe that the quantity $\langle E_{y_1}\cdots E_{y_n}\rangle_X$ must be the joint probability for a sequence $(y_1,\cdots,y_n)$ of n *noninvasive* measurements that couple the same detector to the system n times in succession. Furthermore, the average in Eq. (33) is explicitly different from the nth statistical moment of the raw detector results, $\langle(F_Y)^n\rangle = \sum_{y\in Y}[f_Y(y)]^n P(y)$.

We conclude that, for imperfectly correlated noninvasive detectors, one must perform measurement sequences to obtain the correct statistical moments of an observable using a particular set of CVs. Only for unambiguous measurements with independent probability observables do such measurement sequences reduce to simple powers of the eigenvalues being averaged with single measurement probabilities. If a single measurement by the detector is done per trial, then only the statistical moments of the *detector* observable F_Y can be inferred from that set of CVs, as opposed to the true statistical moments of the inferred system observable F_X.

We can, however, change the CVs to define new observables that correspond to powers of the original observable, such as $G_X = (F_X)^n = \sum_{y\in Y} g_Y(y)E_y$. These new observables can then be measured indirectly using the same experimental setup without the need for measurement sequences. The CVs $g_Y(y)$ for the nth power of F_X will not be a simple power of the CVs $f_Y(y)$ for F_X unless the measurement is unambiguous.

Invasive measurements. If the measurement is invasive, then the disturbance forces us to associate the CVs with the measurement *operations* $\{\mathcal{E}_y\}$ and not solely with their associated probability operators $\{\tilde{E}_y\}$ in order to properly

handle measurement sequences as in Eq. (25). Specifically, we must define the *observable operation*,

$$\mathcal{F}_X = \sum_{y \in Y} f_Y(y) \mathcal{E}_y, \tag{34}$$

which produces the identity $\mathcal{F}_X(1_X) = \sum_{y \in Y} f_Y(y) \tilde{E}_y = F_X$ similar to (32).

Correlated sequences of invasive observable measurements can be obtained by composing the observable operations,

$$\langle (\mathcal{F}_X)^n (1_X) \rangle_X = \sum_{y_1, \dots, y_n} f_Y(y_1) \cdots f_Y(y_n)$$
$$\times \langle \mathcal{E}_{y_1} (\mathcal{E}_{y_2} (\cdots (\tilde{E}_{y_n}) \cdots)) \rangle_X. \tag{35}$$

Such an n-measurement sequence reduces to the nth moment (33) when the disturbance vanishes.

If the time-evolution disturbance $\mathcal{D}_t$ is inserted between different invasive observable measurements, then we obtain an invasive *correlation function* instead,

$$\widetilde{\langle F_X(0) G_X(t) \rangle} = \langle \mathcal{F}_X(\mathcal{D}_t(\mathcal{G}_X(1_X))) \rangle_X. \tag{36}$$

When the observable measurements become noninvasive, then this correctly reduces to the noninvasive correlation function (12). Similarly, n-time invasive correlations can be defined with $n-1$ time-evolution disturbances between the invasive observable measurements $\langle \mathcal{F}_1(\mathcal{D}_{t_1}(\mathcal{F}_2(\cdots \mathcal{D}_{t_{n-1}}(\mathcal{F}_n(1_X)) \cdots))) \rangle$.

Conditioned averages. In addition to statistical moments of the observable, we can also use the CVs to construct principled *conditioned averages* of the observable. Recall that, in the general case of an invasive measurement sequence, we can condition the observable measurement in two distinct ways. If we condition on an outcome z before the measurement of F_X we obtain the *preselected conditioned average* $\langle \widetilde{F_X} \rangle_z$ defined in Eq. (24). On the other hand, if the invasive conditioning measurement of z happens after the invasive observable measurement then we must use the postselected conditional probabilities (27) to construct a *postselected conditioned average*,

$$_z \langle \widetilde{F_X} \rangle = \sum_{y \in Y} f_Y(y) {}_z \langle \tilde{y} \rangle = \frac{\sum_{y \in Y} f_Y(y) \langle \mathcal{E}_y(\tilde{E}'_z) \rangle_X}{\sum_{y \in Y} \langle \mathcal{E}_y(\tilde{E}'_z) \rangle_X}$$
$$= \frac{\langle \mathcal{F}_X(\tilde{E}'_z) \rangle_X}{\langle \mathcal{E}(\tilde{E}'_z) \rangle}. \tag{37}$$

The observable operation $\mathcal{F}_X$ and the nonselective measurement $\mathcal{E}$ encode the relevant details from the first measurement. When the disturbance to the reduced system state vanishes, both the preselected and the postselected conditioned averages simplify to the pure conditioned average $\langle F_X \rangle_z$ defined in Eq. (18) that depends only on the system observable F_X. While the pure conditioned average $\langle F_X \rangle_z$ is independent of the order of conditioning and is always constrained to the eigenvalue range of the observable, the postselected invasive conditioned average $_z \langle \widetilde{F_X} \rangle$ can, perhaps surprisingly, stray outside the eigenvalue range with ambiguous measurements. The combination of the amplified CVs and the disturbance can lead to a postselected average that lies anywhere in the full CV range, rather than just the eigenvalue range. We will see an example of this in Sec. II D 2.

Inversion. So far we have treated the CVs in the expansion (32) as known quantities. However, for a realistic detector situation, the CVs will need to be experimentally determined from the characterization of the detector and the observable that one wishes to measure. The reduced system state P_X will generally not be known *a priori*, since the point of a detector is to learn information about the system in the *absence* of such prior knowledge. We can still solve for the CVs without knowledge of the system state, however, since the probability observables are only specified by the conditional likelihoods $P(y|x)$ that can be obtained independently from detector tomography.

To solve for the CVs when the system state is presumed unknown, we rewrite (32) in the form

$$F_X = \sum_{x \in X} x \sum_{y \in Y} P(y|x) f_Y(y)$$
$$= \sum_{x \in X} x \langle F_Y \rangle_x = \mathcal{S}(F_Y), \tag{38}$$

where $\mathcal{S} = \sum_x x \langle \cdot \rangle_x$ is the map that converts observables in the detector space to observables in the system space $\mathcal{S} : \Sigma_Y^{\mathbb{R}} \to \Sigma_X^{\mathbb{R}}$. Our goal is to invert this map and solve for the required spectrum of F_Y given a desired system observable F_X. However, the inverse of such a map is not generally unique; for it to be uniquely invertible it must be one-to-one between system and detector spaces of equal size. If the detector space is smaller than the system, then no exact inverse solutions are possible; it may be possible, however, to find course-grained solutions that lose some information. Perhaps more alarmingly, if the detector space is larger than the system, then it is possible to have an infinite set of exact solutions.

When disturbance is taken into account as in Eq. (23), the equality (38) becomes

$$F_X = \langle \mathcal{D}(F_Y) \rangle_Y = \mathcal{S}(F_Y), \tag{39}$$

so the composition of the disturbance $\mathcal{D}$ and the detector expectation $\langle \cdot \rangle_Y$ produces the map $\mathcal{S}$ that must be inverted. Equation (38) is a special case when the reduced system state is unchanged by the coupling disturbance.

Pseudoinversion. The entire set of possible solutions to (39) may be completely specified using the *Moore-Penrose pseudoinverse* of the map $\mathcal{S}$, which we denote as $\mathcal{S}^+$. The pseudoinverse is the inverse of the restriction of $\mathcal{S}$ to the space $\Sigma_Y^{\mathbb{R}} \setminus \{F \in \Sigma_Y^{\mathbb{R}} | \mathcal{S}(F) = 0\}$; that is, the null space of $\mathcal{S}$ is removed from the detector space before constructing the inverse. We will show a practical method for computing the pseudoinverse using the singular-value decomposition in the examples to follow.

Using the pseudoinverse, all possible solutions of (39) can be written compactly as

$$F_Y = \mathcal{S}^+(F_X) + (\mathcal{I} - \mathcal{S}^+ \mathcal{S})(G), \tag{40}$$

where $\mathcal{I}$ is the identity map and $G \in \Sigma_Y^{\mathbb{R}}$ is an arbitrary detector observable. The solutions specified by the pseudoinverse in this manner contain exact inverses and course-grainings as special cases.

Detector variance. Since $(\mathcal{I} - \mathcal{S}^+ \mathcal{S})$ is a projection operation to the null space of $\mathcal{S}$, the second term of (40) lives in the

null space of $\mathcal{S}$ and is orthogonal to the first term. Therefore, the norm squared of F_Y has the form

$$||F_Y||^2 = \sum_y [f_Y(y)]^2$$

$$= ||\mathcal{S}^+(F_X)||^2 + ||(\mathcal{I} - \mathcal{S}^+\mathcal{S})(G)||^2, \quad (41)$$

making the $G = 0$ solution have the smallest norm.

The norm $||F_Y||$ of the CV solution is relevant because the second moment of the detector observable F_Y is simply bounded by the norm squared $\langle (F_Y)^2 \rangle = \sum_y P(y)[f_Y(y)]^2 \leqslant ||F_Y||^2$. The second moment is similarly an upper bound for the variance of the detector observable $\mathrm{Var}(F_Y) = \langle (F_Y)^2 \rangle - (\langle F_Y \rangle)^2 \leqslant \langle (F_Y)^2 \rangle$. Therefore, the norm squared is a reasonable upper bound for the detector variance that one can make without prior knowledge of the state.

Mean-squared error. The variance of F_Y governs the mean-squared error of any estimation of its average with a finite sample, such as an empirically measured sample in a laboratory. Specifically, one measures a sequence of detector outcomes of length n, $(y_1, y_2, \ldots, y_n)$ and uses this finite sequence to estimate the average of F_Y via the *unbiased estimator*

$$\overline{F_Y} = \frac{1}{n} \sum_i^n f_Y(y_i) \quad (42)$$

that converges to the true mean value $\langle F_Y \rangle_Y = \langle F_X \rangle$ as $n \to \infty$. The mean-squared error of this estimator $\mathrm{MSE}(\overline{F_Y})$ from the true mean is the variance over the number of trials in the sequence $\mathrm{Var}(F_Y)/n$. Hence, the maximum mean squared error for a finite sequence of length n must be bounded by the norm squared of the CVs divided by length of the sequence,

$$\mathrm{MSE}(\overline{F_Y}) = \frac{\mathrm{Var}(F_Y)}{n} \leqslant \frac{||F_Y||^2}{n}. \quad (43)$$

That is, the norm bounds the number of trials necessary to obtain an experimental estimation of observable averages to a desired precision using the imperfect detector.

Pseudoinverse prescription. Choosing the arbitrary observable to be $G = 0$ therefore not only picks the solution $F_Y = \mathcal{S}^+(F_X)$ that is uniquely related to F_X by discarding the irrelevant null space of $\mathcal{S}$ but also picks the solution with the smallest norm, which places a reasonable upper bound on the statistical error. Without prior knowledge of the system state, the pseudoinverse solution does a reasonable job at obtaining an optimal fit to the relation (39). Moreover, when (39) is not satisfied by the direct pseudoinverse then an exact solution is impossible, but the pseudoinverse still gives the "best fit" course-graining of an exact solution in the least-squares sense. As such, we consider the direct pseudoinverse of F_X to be the preferred solution in the absence of other motivating factors stemming from prior knowledge of the state being measured.

1. Example: Ambiguous marble detector

As an illustrative example similar to the one given in the introduction, suppose that one wishes to know whether the color of a marble is green or red, but one is unable to examine the marble directly. Instead, one only has a machine that can display a blue light or a yellow light after it examines the marble color. In such a case, the marble colors are the propositions of interest, but the machine lights are the only accessible propositions. The lights may be correlated imperfectly with the marble color; for instance, if a blue light is displayed one may learn something about the possible marble color, but it may still be partially *ambiguous* whether the marble is actually green or actually red.

The relevant Boolean algebra for the system is $\Sigma_X = \{0, g, r, 1_X\}$, where g is the proposition for the color green, r is the proposition for the color red, and $1_X = g + r$ is the logical OR of the two possible color propositions. We consider the task of measuring a simple color observable $F_X = (+1)g + (-1)r$ that distinguishes the colors with a sign using an imperfectly correlated detector.

The relevant Boolean algebra for the detector is $\Sigma_Y = \{0, b, y, 1_Y\}$, where b is the proposition for the blue light, y is the proposition for the yellow light, and $1_Y = b + y$. In order to measure the marble observable F_X using only the detector, the experimenter must determine the proper form of the corresponding detector observable F_Y.

First, the experimenter characterizes the detector by sending in known samples and observing the outputs of the detector. After many characterization trials, the experimenter determines to some acceptable precision the four conditional probabilities,

$$P(b|g) = 0.6, \quad P(y|g) = 0.4, \quad (44a)$$

$$P(b|r) = 0.2, \quad P(y|r) = 0.8, \quad (44b)$$

for the detector outcomes b and y given specific marble preparations g and r. These characterization probabilities completely determine the detector response in the form of its *probability observables* (17),

$$E_b = P(b|g)g + P(b|r)r, \quad (45a)$$

$$E_y = P(y|g)g + P(y|r)r. \quad (45b)$$

By construction, $E_b + E_y = g + r = 1_X$.

Second, the experimenter expands the system observable F_X using the detector probability observables Eq. (45) and unknown *contextual values* (CVs) $f_Y(b)$ and $f_Y(y)$ (32),

$$F_X = (+1)g + (-1)r = f_Y(b)E_b + f_Y(y)E_y. \quad (46)$$

After expressing this relation as the equivalent matrix equation,

$$\begin{pmatrix} +1 \\ -1 \end{pmatrix} = \begin{pmatrix} P(b|g) & P(y|g) \\ P(b|r) & P(y|r) \end{pmatrix} \begin{pmatrix} f_Y(b) \\ f_Y(y) \end{pmatrix}, \quad (47)$$

it can be directly inverted to find the CVs (40),

$$f_Y(b) = 3, \quad f_Y(y) = -2. \quad (48)$$

Therefore,

$$F_X = (+1)g + (-1)r = (3)E_b + (-2)E_y, \quad (49)$$

so F_X can be inferred from a measurement of the equivalent detector observable $F_Y = (3)b + (-2)y$.

Notably, the CVs (48) are amplified from the eigenvalues of ± 1 due to the *ambiguity* of the detector. The amplification compensates for the ambiguity so that the correct *average* can be obtained after measuring an ensemble of many

unknown marbles described by the initial marble state P_X. The amplification also leads to a larger upper bound for the variance (41) of the detector,

$$||F_Y||^2 = 13. \tag{50}$$

Hence, we can expect the imperfect detector to display a root-mean-square (rms) error (43) in the reported average color that is no larger than $\sqrt{13/n} \approx 3.6/\sqrt{n}$ after n repeated measurements. For contrast, a perfect detector would display an rms error no larger than $\sqrt{2/n} \approx 1.4/\sqrt{n}$ after n repeated measurements.

2. Example: Invasive ambiguous detector

The detector apparatus in the last example could be generally *invasive*. In such a case, the characterization probabilities (44) composing the probability observables (45) would be a combination of the initial state of the detector lights P_Y and a *disturbance* $\mathcal{D}$ from the measurement coupling according to (23),

$$\tilde{P}(b|g) = P_Y(b)[D_{g,b}(gb) + D_{g,b}(rb)] \\ + P_Y(y)[D_{g,y}(gb) + D_{g,y}(rb)], \tag{51a}$$

$$\tilde{P}(y|g) = P_Y(b)[D_{g,b}(gy) + D_{g,b}(ry)] \\ + P_Y(y)[D_{g,y}(gy) + D_{g,y}(ry)], \tag{51b}$$

$$\tilde{P}(b|r) = P_Y(b)[D_{r,b}(gb) + D_{r,b}(rb)] \\ + P_Y(y)[D_{r,y}(gb) + D_{r,y}(rb)], \tag{51c}$$

$$\tilde{P}(y|r) = P_Y(b)[D_{r,b}(gy) + D_{r,b}(ry)] \\ + P_Y(y)[D_{r,y}(gy) + D_{r,y}(ry)], \tag{51d}$$

where we have used the marginalization identity $D_{c,d}(b) = D_{c,d}(gb) + D_{c,d}(rb)$ for $c \in \{g,r\}$ and $d \in \{b,y\}$. For a noninvasive detector, the transition probabilities that involve marbles changing color must be zero $D_{g,b}(rb) = D_{g,b}(ry) = D_{g,y}(ry) = D_{g,y}(rb) = D_{r,b}(gb) = D_{r,b}(gy) = D_{r,y}(gb) = D_{r,y}(gy) = 0$. However, they need not be zero for a general invasive detector.

As an example, suppose that the initial detector state is unbiased, $P_Y(b) = P_Y(y) = 1/2$, and that the detector has a 10% chance of flipping the color of a given marble. The following possible values for the sixteen transition probabilities would then lead to the same effective characterization probabilities (44) as before,

$$D_{g,b}(gb) = 0.5, \quad D_{g,y}(gb) = 0.5, \tag{52a}$$

$$D_{g,b}(gy) = 0.3, \quad D_{g,y}(gy) = 0.3, \tag{52b}$$

$$D_{r,b}(rb) = 0.1, \quad D_{r,y}(rb) = 0.1, \tag{52c}$$

$$D_{r,b}(ry) = 0.7, \quad D_{r,y}(ry) = 0.7, \tag{52d}$$

$$D_{g,b}(rb) = 0.1, \quad D_{g,y}(rb) = 0.1, \tag{52e}$$

$$D_{g,b}(ry) = 0.1, \quad D_{g,y}(ry) = 0.1, \tag{52f}$$

$$D_{r,b}(gb) = 0.1, \quad D_{r,y}(gb) = 0.1, \tag{52g}$$

$$D_{r,b}(gy) = 0.1, \quad D_{r,y}(gy) = 0.1. \tag{52h}$$

Since the effective characterization probabilities are the same, the probability observables are the same as (45), leading to the same CVs as (48) to measure the observable $F_X = (+1)g + (-1)r$.

The disturbance of the reduced marble state will become apparent only when making a second measurement after the first one. Suppose we make a second measurement of the marble colors g and r directly. The probability of obtaining a detector outcome $d \in \{b,y\}$ and then observing a specific marble color $c \in \{g,r\}$ will then be $P_X(g)[P_Y(b)D_{g,b}(cd) + P_Y(y)D_{g,y}(cd)] + P_X(r)[P_Y(b)D_{r,b}(cd) + P_Y(y)D_{r,y}(cd)]$. If we define an *operation* as in Eq. (25) to be

$$\mathcal{E}_d(c) = \langle \mathcal{D}(cd) \rangle_Y = g[P_Y(b)D_{g,b}(cd) + P_Y(y)D_{g,y}(cd)] \\ + r[P_Y(b)D_{r,b}(cd) + P_Y(y)D_{r,y}(cd)], \tag{53}$$

then we can express the probability for the sequence compactly as $\langle \mathcal{E}_d(c) \rangle_X$.

Averaging the outcomes for the detector lights using the CVs (48) and then conditioning on a particular marble color c in the second measurement produces a *postselected conditioned average* of the marble colors (37) as reported by the invasive ambiguous detector,

$$_c\langle \widetilde{F_X} \rangle = \frac{f_Y(b)\langle \mathcal{E}_b(c) \rangle_X + f_Y(y)\langle \mathcal{E}_y(c) \rangle_X}{\langle \mathcal{E}_b(c) \rangle_X + \langle \mathcal{E}_y(c) \rangle_X}. \tag{54}$$

If we also preselect the marbles to be a particular color, we can compute the pre- and postselected conditioned averages of the marble colors as reported by the invasive ambiguous detector from (48), (52), and (54),

$$_g\langle \widetilde{F_X} \rangle_g = 1.125, \tag{55a}$$

$$_r\langle \widetilde{F_X} \rangle_g = 0.5, \tag{55b}$$

$$_g\langle \widetilde{F_X} \rangle_r = 0.5, \tag{55c}$$

$$_r\langle \widetilde{F_X} \rangle_r = -1.375. \tag{55d}$$

Due to a combination of the invasiveness and the ambiguity of the measurement, the postselected conditioned averages can stray outside the eigenvalue range $[-1,1]$ for the observable F_X. However, they remain within the CV range $[-2,3]$. When the measurement is noninvasive, then the pre- and postselected conditioned averages in Eq. (55) that remain well-defined reduce to the pure conditioned averages $\langle F_X \rangle_g = 1$ and $\langle F_X \rangle_r = -1$.

3. Example: Redundant ambiguous detector

Consider a similar marble detection setup to the previous examples, but where the detector apparatus has *three* independent outcome lights: blue, yellow, and purple. Hence, the detector Boolean algebra is $\Sigma_Y = \{0, b, y, p, b + y, b + p, y + p, 1_Y\}$, where p is the new proposition for the purple light, and $1_Y = b + y + p$. After characterizing the detector the experimenter finds the conditional probabilities

$$P(b|g) = 0.5, \quad P(y|g) = 0.3, \quad P(p|g) = 0.2, \tag{56a}$$

$$P(b|r) = 0.1, \quad P(y|r) = 0.7, \quad P(p|r) = 0.2 \tag{56b}$$

that define the probability observables,

$$E_b = P(b|g)g + P(b|r)r, \tag{57a}$$

$$E_y = P(y|g)g + P(y|r)r, \tag{57b}$$

$$E_p = P(p|g)g + P(p|r)r. \tag{57c}$$

By construction, $E_b + E_y + E_p = 1_X$. Furthermore, $E_p = (0.2)1_X$, so the purple outcome cannot distinguish whether the marble is green or red and can be imagined as a generic detector malfunction outcome.

The experimenter now has a choice for how to assign CVs to a detector observable F_Y in order to infer the marble observable $F_X = (+1)g + (-1)r$. A simple choice is to ignore the redundant (and nondistinguishing) purple outcome by zeroing out its CV $f_Y(p) = 0$, and then invert the remaining relationship analogously to (47) to find $f_Y(b) = 3.125$ and $f_Y(y) = -1.875$. The variance bound for this simple choice is $||F_Y||^2 = 13.2813$, leading to a root-mean-square error no larger than $\sqrt{13.2813/n} \approx 3.6/\sqrt{n}$ after n repeated measurements.

However, a better choice is to find the preferred values for all three outcomes using the pseudoinverse (40) of the map between F_Y and F_X. To do this, we write a matrix equation similar to (47) that uses all three outcomes:

$$\begin{pmatrix} +1 \\ -1 \end{pmatrix} = \mathcal{S} \begin{pmatrix} f_Y(b) \\ f_Y(y) \end{pmatrix}, \tag{58a}$$

$$\mathcal{S} = \begin{pmatrix} P(b|g) & P(y|g) & P(p|g) \\ P(b|r) & P(y|r) & P(p|r) \end{pmatrix}. \tag{58b}$$

The pseudoinverse $\mathcal{S}^+$ can be constructed by using the singular-value decomposition, $\mathcal{S} = \mathcal{U}\Sigma\mathcal{V}^T$, where $\mathcal{U}$ is an orthogonal matrix composed of the normalized eigenvectors of $\mathcal{S}\mathcal{S}^T$, $\mathcal{V}$ is an orthogonal matrix composed of the normalized eigenvectors of $\mathcal{S}^T\mathcal{S}$, and Σ is a diagonal matrix composed of the singular values of $\mathcal{S}$ (which are the square roots of the eigenvalues of $\mathcal{S}\mathcal{S}^T$ and $\mathcal{S}^T\mathcal{S}$). After computing the singular-value decomposition, the pseudoinverse can be constructed as $\mathcal{S}^+ = \mathcal{V}\Sigma^+\mathcal{U}^T$, where Σ^+ is the diagonal matrix constructed by inverting all nonzero elements of Σ^T. Performing this inversion, we find the following preferred CV:

$$\mathcal{S}^+ = \frac{5}{36} \begin{pmatrix} 15 & -7 \\ -3 & 11 \\ 3 & 1 \end{pmatrix}, \tag{59a}$$

$$\begin{pmatrix} f_Y(b) \\ f_Y(y) \\ f_Y(p) \end{pmatrix} = \mathcal{S}^+ \begin{pmatrix} +1 \\ -1 \end{pmatrix} = \frac{5}{18} \begin{pmatrix} 11 \\ -7 \\ 1 \end{pmatrix} = \begin{pmatrix} 3.0\bar{5} \\ -1.9\bar{4} \\ 0.2\bar{7} \end{pmatrix}. \tag{59b}$$

This preferred solution has the smallest variance bound of $||F_Y||^2 = 13.1944$.

We find (perhaps counterintuitively) that, even though the purple outcome itself cannot distinguish the marble color, the fact that one obtains a purple outcome at all still provides some useful information to the experimenter due to the asymmetry of the blue and yellow outcomes. Indeed, if for the red marble we instead found the symmetric detector response $P(b|r) = 0.3$, $P(y|r) = 0.5$, and $P(p|r) = 0.2$, the pseudoinverse would produce the preferred CV $f_Y(b) = 5$, $f_Y(y) = -5$, and $f_Y(p) = 0$, indicating that the purple outcome was truly noninformative.

A less principled approach to solving (58) would be for the experimenter to assign a completely arbitrary value to one outcome, like $f_Y(b) = B$. The CV relation still produces a matrix equation,

$$\begin{pmatrix} +1 - BP(b|g) \\ -1 - BP(b|r) \end{pmatrix} = \begin{pmatrix} P(y|g) & P(p|g) \\ P(y|r) & P(p|r) \end{pmatrix} \begin{pmatrix} f_Y(y) \\ f_Y(p) \end{pmatrix}, \tag{60}$$

that can be solved to find

$$f_Y(y) = B - 5, \quad f_Y(p) = 12.5 - 4B. \tag{61}$$

The bound for the variance of this solution is $||F_Y||^2 = 18B^2 - 110B + 181.25 \geqslant 13.1944$; the value of B that minimizes the bound is $B = 3.0\bar{5}$, which recovers the pseudoinverse solution.

Although picking an arbitrary solution gives mathematically equivalent results, the experimenter will only increase the norm of the solution without any physical motivation. As such, the higher moments of the detector observable F_Y can be correspondingly larger, and more trials may be necessary for the estimated average of the system observable F_X to reach the desired precision.

4. Example: Continuous detector

Consider the extreme example of a marble color detector that has a continuum of outcomes, such as the position of impact of a marble on a continuous screen. In such a case, the detector sample space Y is indexed by a real parameter $y \in \mathbb{R}$, and the relevant Boolean algebra Σ_Y can be chosen to be the set of all Borel subsets of the real line [75,76].

After characterizing the detector, the experimenter finds that the detector displaces its initial probability distribution $dP_Y(y) = p_Y(y)dy$ by an amount z from the zero-point according to which marble-color is sent into the detector,

$$dP(y|g) = dP_Y(y - z), \quad dP(y|r) = dP_Y(y + z). \tag{62}$$

These probabilities define the probability observables,

$$dE(y) = g\,dP(y|g) + r\,dP(y|r), \tag{63}$$

such that $\int_{\mathbb{R}} dE(y) = 1_X$.

To infer information about the marble observable F_X using this detector, the experimenter must assign a continuum of CV $f_Y(y)$ such that

$$F_X = (+1)g + (-1)r = \int_{\mathbb{R}} f_Y(y)dE(y) \tag{64}$$

or, in matrix form,

$$\begin{pmatrix} +1 \\ -1 \end{pmatrix} = \mathcal{S}[f_Y] = \begin{pmatrix} \int_{\mathbb{R}} f_Y(y)dP_Y(y - z) \\ \int_{\mathbb{R}} f_Y(y)dP_Y(y + z) \end{pmatrix}. \tag{65}$$

Since f_Y is a function, $\mathcal{S}$ is a vector-valued functional, which is why we adopt the square-bracket notation.

In this case, the detector outcomes are overwhelmingly redundant. However, we can pick the least norm solution using the pseudoinverse of the map $\mathcal{S}$ as before. To do so, we first calculate $\mathcal{S}\mathcal{S}^T$:

$$\mathcal{S}^T = \begin{pmatrix} p_Y(y - z) & p_Y(y + z) \end{pmatrix}, \tag{66a}$$

$$\mathcal{S}\mathcal{S}^T = \begin{pmatrix} a & b(z) \\ b(z) & a \end{pmatrix}, \tag{66b}$$

where

$$a = \int_{\mathbb{R}} p_Y(y) dP_Y(y) = \int_{\mathbb{R}} p_Y^2(y) dy, \qquad (67a)$$

$$b(z) = \int_{\mathbb{R}} p_Y(y+z) p_Y(y-z) dy, \qquad (67b)$$

and we find its eigenvalues of $a + b(z)$ with corresponding normalized eigenvector $(1,1)/\sqrt{2}$ and $a - b(z)$ with corresponding normalized eigenvector $(-1,1)/\sqrt{2}$. We can then construct the orthogonal matrix $\mathcal{U}$ composed of the normalized eigenvectors of $\mathcal{S}\mathcal{S}^T$ and the diagonal matrix Σ composed of the square roots of the eigenvalues of $\mathcal{S}\mathcal{S}^T$,

$$\mathcal{U} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}, \qquad (68)$$

$$\Sigma = \begin{pmatrix} \sqrt{a + b(z)} & 0 \\ 0 & \sqrt{a - b(z)} \end{pmatrix}. \qquad (69)$$

Next we calculate the relevant eigenfunctions of $\mathcal{S}^T \mathcal{S}$ that correspond to the same nonzero eigenvalues $a \pm b(z)$ of $\mathcal{S}\mathcal{S}^T$; the remaining eigenfunctions belong to the null space of $\mathcal{S}$ and

do not contribute. Specifically, we have

$$\mathcal{S}^T \mathcal{S}[h](y) = p_Y(y-z) \int_{\mathbb{R}} h(y) dP_Y(y-z)$$
$$+ p_Y(y+z) \int_{\mathbb{R}} h(y) dP_Y(y+z), \quad (70)$$

where h is an arbitrary function. Then the equations

$$\mathcal{S}^T \mathcal{S}[v_+](y) = [a + b(z)]v_+(y), \qquad (71a)$$
$$\mathcal{S}^T \mathcal{S}[v_-](y) = [a - b(z)]v_-(y), \qquad (71b)$$

define the normalized eigenfunctions

$$v_+(y) = \frac{p_Y(y-z) + p_Y(y+z)}{\sqrt{2[a + b(z)]}}, \qquad (72a)$$

$$v_-(y) = -\frac{p_Y(y-z) - p_Y(y+z)}{\sqrt{2[a - b(z)]}}, \qquad (72b)$$

which allows us to construct the relevant part of the orthogonal map $\mathcal{V}^T$,

$$\mathcal{V}^T[h] = \left(\int v_+(y) h(y) dy \quad \int v_-(y) h(y) dy \right), \quad (73)$$

completing the nonzero part of the singular-value decomposition of $\mathcal{S} = \mathcal{U}\Sigma\mathcal{V}^T$.

Finally, we construct the pseudoinverse,

$$\mathcal{S}^+ = \mathcal{V}\Sigma^+\mathcal{U}^T = \left(\frac{v_+(y)}{\sqrt{2[a + b(z)]}} - \frac{v_-(y)}{\sqrt{2[a - b(z)]}} \quad \frac{v_+(y)}{\sqrt{2[a + b(z)]}} + \frac{v_-(y)}{\sqrt{2[a - b(z)]}} \right), \qquad (74)$$

and solve for the CV,

$$f_Y(y) = \mathcal{S}^+ \begin{pmatrix} +1 \\ -1 \end{pmatrix} = \frac{p_Y(y-z) - p_Y(y+z)}{a - b(z)}, \qquad (75)$$

where a and $b(z)$ are as defined in Eq. (67).

The pseudoinverse solution (75) contains only the physically relevant detector state density p_Y and provides direct physical intuition about the detection process. Namely, everything in the shifted distribution corresponding to the green marble $p_Y(y-z)$ is associated with the eigenvalue $+1$, while everything in the shifted distribution corresponding to the red marble $p_Y(y+z)$ is associated with the eigenvalue -1. The overall amplification factor $a - b(z)$ indicates the discrepancy between the overlap of the shifted distributions and the distribution autocorrelation. The more the shifted distributions overlap, the more ambiguous the measurement will be, so the amplification factor makes the CVs larger to compensate. If the shifted distributions do not overlap, then $b(z) \to 0$ and the only amplification comes from the autocorrelation a that indicates the ambiguity of the intrinsic profile of the detector state. Moreover, the support of the CVs is equal to the support of both shifted detector distributions, which is physically satisfying.

The bound for the detector variance using the pseudoinverse solution is $||f_Y||^2 = 2/[a - b(z)]$, which depends solely on the amplification factor in the denominator. If the measurement is strong, such that $a - b(z) = 1$, then the variance bound

reduces to the ideal variance bound of 2, as expected, leading to a maximum rms error of $\sqrt{2/n}$. Any additional ambiguity of the measurement stemming from distribution overlap or distributed autocorrelation amplifies the maximum rms error by a factor of $\sqrt{1/[a - b(z)]}$.

Contrast these preferred values with the generic linear solution $f_Y(y) = y/z$, which also satisfies (64) when p_Y is symmetric about its mean [17,23,49]. While the generic solution could be argued to be simpler in form, it provides no information about the detector and provides no physical insight into the meaning or origin of the values themselves. It has nonzero support in areas where the detector has zero support and even gets progressively larger in regions that will not contribute to the average. Moreover, the bound for the detector variance diverges, indicating that the rms error can in principle be unbounded. Hence, despite the mathematical equivalence, the linear solution is *physically* inferior as a solution when compared to the pseudoinverse (75).

III. QUANTUM PROBABILITY THEORY

To transition from the classical theory of probability to the quantum theory we shall take a somewhat unconventional approach that leverages what we have already derived in the classical theory. Specifically, we shall construct the quantum theory as a superstructure over the existing classical theory, rather than developing it as an independent logical system

[2,70] or as a restriction of a larger classical theory [64,65]. This approach serves to illustrate the myriad similarities between the quantum and classical theories, while also highlighting their key differences. We shall see that the contextual-value formalism is essentially unchanged, despite the modifications that must be made to the operational theory of measurement.

A. Sample spaces and observables

Quantum sample space. The quantum theory of probability forms a superstructure on the classical theory of probability in the following sense: given a classical sample space X, the corresponding quantum sample space can be obtained as the orbit of X under the action of the special unitary group of rotations. That is, the entire classical sample space X can be rotated to a different classical sample space $X' = \mathcal{U}(X)$ with some special unitary rotation $\mathcal{U}$. We call each classical sample space generated in this fashion a *framework* to be consistent with other recent work [77]. The collection of all such continuously connected classical sample spaces is the quantum sample space, which we will notate as $\mathcal{Q}(X)$ to emphasize that it can be generated from X.

Representation. If the sample space X is represented as a set of orthogonal rank-1 projections $\{|x\rangle\langle x|\}$ on a Hilbert space, the rotated sample space $X' = \mathcal{U}(X)$ will be represented by a different set of orthogonal projections $\{\mathcal{U}(|x\rangle\langle x|)\}$ on the same Hilbert space. Any such rotation $\mathcal{U}$ can be given a spinor representation (see, e.g., [78–80]) as a two-sided product with a rotor U belonging to the special unitary group, such that $U^\dagger U = U U^\dagger = 1_X$, and $(U^\dagger)^\dagger = U$. The *involution* (†) is the adjoint with respect to the inner product of the Hilbert space. While the projections $\{|x\rangle\langle x|\}$ correspond to subspaces spanned by vectors $\{|x\rangle\}$ in the Hilbert space, the rotated projections $\{U^\dagger|x\rangle\langle x|U\}$ correspond to subspaces spanned by rotated vectors $\{U^\dagger|x\rangle\}$. In what follows we shall tend to use the shorter algebraic notation x and adopt the equivalent Hilbert space notation $|x\rangle\langle x|$ only when it readily simplifies expressions.

Since the Hilbert space representation of a unitary rotor U generally contains complex numbers in order to satisfy the special unitary group relations, the Hilbert space also becomes *complex*. However, it is important to note that the complex structure arises solely from the representation of the unitary rotations that specify the relative framework orientations and will not appear directly in any calculable quantity to follow [81]. The representation of the quantum sample space $\mathcal{Q}(X)$ therefore consists of all possible rank-1 projections on the complex Hilbert space in which the classical sample space X is represented.

Quantum observables. Each classical framework X has an associated Boolean algebra Σ_X and space of observables $\Sigma_X^\mathbb{R}$ exactly as previously discussed. The space of quantum observables is the collection of all classical observables that are independently constructed in all the classical frameworks in $Q(X)$. We will denote this space as $\Sigma_{\mathcal{Q}(X)}^\mathbb{R}$. Quantum observables are therefore constructed entirely with real numbers that have empirical meaning for a laboratory setting; hence, their representations on a complex Hilbert space will be Hermitian operators.

For observables in the same framework $A, B \in \Sigma_X^\mathbb{R}$, we find that $\mathcal{U}(A)\mathcal{U}(B) = U^\dagger A U U^\dagger B U = U^\dagger A B U = \mathcal{U}(AB)$, meaning that the rotations preserve their algebraic structure. As a corollary, all observables in $\Sigma_{\mathcal{Q}(X)}^\mathbb{R}$ can be obtained by rotating observables constructed in a single framework $\Sigma_X^\mathbb{R}$; hence, our previous discussion of observables carries over to the quantum theory essentially unaltered.

Furthermore, the independence of the propositions in a framework X remains unaltered by unitary rotation, so every other framework X' has the same number of independent propositions. Thus, the number of independent propositions is an invariant known as the *quantum dimension*; for a representation it fixes the dimension of the Hilbert space. Similarly, the identity and zero observables are invariants, so are the same in every framework and unique in the quantum observable algebra.

Since each different framework forms a separate well-behaved classical sample space, the entire preceding discussion about classical probability theory applies unaltered when restricted to a particular framework in the quantum theory. All observables constructed in a particular framework will commute with each other. We expect distinctly quantum features to appear only when comparing elements from different frameworks.

Noncommutativity. The unitary rotations $\mathcal{U}$ are generally *noncommutative* and so introduce noncommutativity into the quantum theory that is not present in the classical theory. Specifically, given $A, B \in \Sigma_X^\mathbb{R}$, $A' = \mathcal{U}(A)$, and $B' = \mathcal{V}(B)$, then $A'B' = U^\dagger A U V^\dagger B V \neq B'A'$, since U and V do not necessarily commute with each other or with A and B. Such noncommutativity is a manifestation of the fact that the Boolean algebras corresponding to different frameworks are *incompatible* with each other; propositions from one framework cannot form a Boolean logical AND with propositions from a different framework. We shall see in the next section, however, that the notion of *disturbance* followed by a logical AND can be generalized to the noncommutative setting in the form of the projection postulate.

Disturbance. All nonconditioning disturbances $\mathcal{D}$ in the quantum theory also take the form of unitary rotations $\mathcal{U}$. Indeed, we shall see that the parallels between the quantum theory and the classical theory with disturbance are quite strong when one interprets all unitary rotations as a form of classical disturbance.

Time evolution. As an example, the continuous time-evolution of a closed quantum system is specified by a disturbance in the form of a unitary rotation $\mathcal{U}_t$ with corresponding rotor U_t, known as a *propagator*. For nonrelativistic quantum mechanics, the time-dependence of the rotor is specified by the Schrödinger equation: $\partial_t U_t = (H/i\hbar)U_t$ and a Hamiltonian observable H that generates the time translation. We are not concerned with the (well-established) details of continuous time-evolution in this paper, so we will treat any unitary rotations as given in what follows.

1. Example: Polarization

As an example quantum system we shall pick the simplest possible nontrivial system: a qubit. Specifically, we will consider the polarization degree of freedom of a laser beam.

Suppose we are interested in measuring the linear polarization of the beam with respect to the surface of an optical table. We denote the polarization direction parallel to the table as "horizontal" (h) and the direction perpendicular to the table as "vertical" (v). Although we casually refer to the polarizations h and v as if they were properties of the light beam, the propositions h and v operationally refer to two independent outcomes of a polarization distinguishing device, such as a polarizing beam splitter, that can be implemented in the laboratory. The two orthogonal polarizations form a classical sample space $X = \{h, v\}$ and a classical Boolean algebra $\Sigma_X = \{0, h, v, 1_X\}$, where $1_X = h + v$, similar to the classical sample space for the marble colors considered in Sec. II D 1. By extending the Boolean algebra over the reals to $\Sigma_X^{\mathbb{R}}$ as before we can define classical observables $F_X = ah + bv$ in this sample space, such as the Stokes observable $S_X = h - v$ that distinguishes the polarizations with a sign.

We can represent the commutative observable algebra $\Sigma_X^{\mathbb{R}}$ as diagonal 2×2 matrices,

$$h = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad v = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad F_X = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}, \quad (76)$$

which can also be understood as commuting Hermitian operators over a two-dimensional Hilbert space. The atomic propositions $h = |h\rangle\langle h|$ and $v = |v\rangle\langle v|$ are projectors that correspond to disjoint subspaces spanned by the orthonormal Jones' polarization basis for the Hilbert space,

$$|h\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |v\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (77)$$

To obtain the full quantum sample space $Q(X)$ from X, we introduce the group of possible polarization rotations. Algebraically, an arbitrary rotation $\mathcal{U}(F_X) = U^\dagger F_X U$ can be readily understood in terms of its rotor U, which is an element of the group SU(2) and can be parametrized, for example, in terms of the Cartan decomposition $U_{\alpha,\beta,\gamma} = \exp(i\alpha\sigma_z/2)\exp(i\beta\sigma_y/2)\exp(i\gamma\sigma_z/2)$, which for a qubit happens to correspond to an Euler angle decomposition of a three-dimensional rotation. Here $i\sigma_z$ and $-i\sigma_y$ are two of the three generators of the Lie algebra for SU(2) in terms of the standard Pauli matrices:

$$\sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (78)$$

Since the group generators have a complex representation, the unitary rotation $U_{\alpha,\beta,\gamma}$ will also have a complex representation in the Hilbert space,

$$e^{i\frac{\alpha}{2}\sigma_z} = \begin{pmatrix} e^{i\frac{\alpha}{2}} & 0 \\ 0 & e^{-i\frac{\alpha}{2}} \end{pmatrix}, \quad (79a)$$

$$e^{i\frac{\beta}{2}\sigma_y} = \begin{pmatrix} \cos\frac{\beta}{2} & \sin\frac{\beta}{2} \\ -\sin\frac{\beta}{2} & \cos\frac{\beta}{2} \end{pmatrix}, \quad (79b)$$

$$U_{\alpha,\beta,\gamma} = \begin{pmatrix} e^{i(\alpha+\gamma)/2}\cos\frac{\beta}{2} & e^{i(\alpha-\gamma)/2}\sin\frac{\beta}{2} \\ -e^{-(\alpha-\gamma)/2}\sin\frac{\beta}{2} & e^{-i(\alpha+\gamma)/2}\cos\frac{\beta}{2} \end{pmatrix}. \quad (79c)$$

The algebraic involution $U_{\alpha,\beta,\gamma}^\dagger$ is the complex transpose in the matrix representation.

Physically, the factor $\exp(i\beta\sigma_y/2)$ corresponds to a rotation of the apparatus around the axis of the light beam by an

angle $\beta/2$, while the factors $\exp(i\alpha\sigma_z/2)$ and $\exp(i\gamma\sigma_z/2)$ correspond to the action of phase plates that shift the relative phases of h and v by $\alpha/2$ and $\gamma/2$, respectively. Hence, the ubiquitous quantum phase also appears as a consequence of the unitary rotations.

Using the unitary rotations, we can generate other incompatible frameworks $\mathcal{U}_{\alpha,\beta,\gamma}(X) = \{\mathcal{U}_{\alpha,\beta,\gamma}(h), \mathcal{U}_{\alpha,\beta,\gamma}(v)\}$ in $Q(X)$,

$$\mathcal{U}_{\alpha,\beta,\gamma}(h) = U_{\alpha,\beta,\gamma}^\dagger h U_{\alpha,\beta,\gamma} = \begin{pmatrix} \cos^2\frac{\beta}{2} & \frac{1}{2}e^{-i\gamma}\sin\beta \\ \frac{1}{2}e^{i\gamma}\sin\beta & \sin^2\frac{\beta}{2} \end{pmatrix}, \quad (80a)$$

$$\mathcal{U}_{\alpha,\beta,\gamma}(v) = U_{\alpha,\beta,\gamma}^\dagger v U_{\alpha,\beta,\gamma} = \begin{pmatrix} \sin^2\frac{\beta}{2} & -\frac{1}{2}e^{-i\gamma}\sin\beta \\ -\frac{1}{2}e^{i\gamma}\sin\beta & \cos^2\frac{\beta}{2} \end{pmatrix}, \quad (80b)$$

which depend solely on the two parameters β and γ. The atomic propositions of such a rotated framework are projectors corresponding to each disjoint subspace spanned by a rotated orthonormal Jones' polarization basis,

$$U_{\alpha,\beta,\gamma}^\dagger |h\rangle = \begin{pmatrix} e^{-i(\alpha+\gamma)/2}\cos\frac{\beta}{2} \\ e^{-i(\alpha-\gamma)/2}\sin\frac{\beta}{2} \end{pmatrix}, \quad (81a)$$

$$U_{\alpha,\beta,\gamma}^\dagger |v\rangle = \begin{pmatrix} -e^{i(\alpha-\gamma)/2}\sin\frac{\beta}{2} \\ e^{i(\alpha+\gamma)/2}\cos\frac{\beta}{2} \end{pmatrix}. \quad (81b)$$

Physically, one could in principle construct an apparatus corresponding to such a rotated framework using three laboratory elements: (1) attach a tunable phase plate to the incident port of a polarizing beam splitter with the fast axis aligned to the table, (2) rotate both the beam splitter and attached phase plate with respect to the table, and (3) attach a second tunable phase plate to the incident port of the first phase plate with the fast axis aligned to the table. Of course this is only one possible parametrization for the unitary rotations; other parametrizations will correspond to other experimental implementations.

It follows that any observable in the full quantum observable space $\Sigma_{Q(X)}^{\mathbb{R}}$ can be obtained by rotating a classical observable $F_X = ah + bv$ to the appropriate framework,

$$F_{X'} = \mathcal{U}_{\alpha,\beta,\gamma}(F_X) = a\mathcal{U}_{\alpha,\beta,\gamma}(h) + b\mathcal{U}_{\alpha,\beta,\gamma}(v),$$

$$= \begin{pmatrix} \frac{a+b}{2} + \frac{a-b}{2}\cos\beta & \frac{a-b}{2}e^{-i\gamma}\sin\beta \\ \frac{a-b}{2}e^{i\gamma}\sin\beta & \frac{a+b}{2} - \frac{a-b}{2}\cos\beta \end{pmatrix}. \quad (82)$$

We see that a general qubit observable depends on four parameters: the eigenvalues a and b, as well as the framework orientation angles β and γ. The complex representation of an observable stems solely from the unitary rotation of the atomic propositions h and v to a different relative framework. The observables no longer generally commute since the unitary rotations need not commute.

B. States, densities, and collapse

Quantum states. A quantum state P is a classical state defined in a particular framework X that is then extended to apply to the entire quantum Boolean algebra $\Sigma_{Q(X)}$. The

extension of a classical state P that has been defined in a framework X to a proposition $x' = \mathcal{U}(x) \in X' = \mathcal{U}(X)$ in a different framework can be accomplished by heuristically breaking down the state into a composition of the classical state in framework X and *transition probabilities* $D_x(x')$ that connect the framework X to the different framework X',

$$P(x') = \sum_{x \in X} P(x) D_x(x'). \tag{83}$$

The transition probabilities characterize a *disturbance* (II B) that connects the classical state P to propositions in incompatible frameworks.

To define the transition probabilities, we assume that atomic propositions in the framework X are undisturbed, so $D_x(x) = 1$. The only classical state with this property is the pure state which has a trace-density (6) $\rho = x$. Hence, we assume that we can consistently write the transition probability $D_x(x')$ in terms of the extension of the *trace* to the full Boolean algebra $\Sigma_{\mathcal{Q}(X)}$,

$$D_x(x') = \text{Tr}(x x'). \tag{84}$$

Notably, this definition makes the transition between frameworks symmetric.

Born rule. We pick the trace extension to be the unique measure that satisfies the cyclic property $\text{Tr}(AB) = \text{Tr}(BA)$ for all $A, B \in \Sigma_{\mathcal{Q}(X)}$ and agrees with the classical trace (7) within any specific framework [82]. On a Hilbert space, (84) has the familiar form

$$D_x(x') = \text{Tr}(|x\rangle\langle x||x'\rangle\langle x'|) = |\langle x|x'\rangle|^2, \tag{85}$$

which we immediately recognize as the *Born rule* [83]. Hence, the complex square of the Hilbert space inner product can be seen as a disguised form of the natural extension of the trace to define transition probabilities between propositions in incompatible frameworks. If we recall that $x' = \mathcal{U}(x) = U^\dagger x U$ we can also write the transition probability (85) in terms of the unitary rotor that connects the two propositions, $D_x(x') = \text{Tr}(|x\rangle\langle x|U^\dagger|x\rangle\langle x|U) = |\langle x|U|x\rangle|^2$.

Density operator. We can rewrite (83) in a more familiar form by using the Born rule (84) and the full trace density (6) of the original state $\rho = \sum_{x \in X} P(x) x$, which is traditionally known as the *density operator*,

$$P(x') = \sum_{x \in X} P(x) \text{Tr}(x x') = \text{Tr}(\rho x'). \tag{86}$$

This form of the probability functional conforms to *Gleason's theorem* [84]. We note, however, that it is the extension of the *trace* that extends the state to the noncommutative quantum setting since the trace-density ρ is identical to a classical trace density in some particular framework X.

Moments. Since the probabilities $P(x')$ are well defined for a proposition in any framework $x' \in X'$, we can linearly extend P to an expectation functional $\langle \cdot \rangle$ on the entire quantum observable algebra $\Sigma_{\mathcal{Q}(X)}^{\mathbb{R}}$,

$$\langle F_{X'} \rangle = \sum_{x' \in X'} f_{X'}(x') P(x') = \text{Tr}(\rho F_{X'}). \tag{87}$$

Similarly, observable moments will be well defined by the expectation functional

$$\langle (F_{X'})^n \rangle = \sum_{x' \in X'} f_{X'}^n(x') P(x') = \text{Tr}[\rho(F_{X'})^n]. \tag{88}$$

Hence, the unitary rotations and resulting extension of the trace completely construct the quantum probability space from a single classical probability space and its associated observables.

Double-sided AND. To be consistent with the assumptions made in Eq. (84), we must also ensure that conditioning a quantum state on an atomic proposition will collapse the state to a pure state with a trace density equal to that atomic proposition. In other words, we must generalize the logical AND of the classical case to the noncommutative incompatible frameworks in the quantum case. The consistent way to do this is through a *double-sided product*: given atomic propositions $x \in X$ and $x' \in X'$ then $x' x x' = |x'\rangle\langle x'|x\rangle\langle x|x'\rangle\langle x'| = \text{Tr}(x x') x' = D_x(x') x'$.

The double-sided product with x' produces a *transition probability* $D_x(x')$ from x to x' as a proportionality factor in addition to collapsing the original proposition x to x'. In this sense, the double-sided product includes a form of *disturbance* in addition to the logical *and* of pure conditioning. If $X = X'$, so the frameworks coincide, then x and x' will commute; the disturbance will vanish, reducing the transition probability $D_x(x')$ to either 0 or 1; and, the classical AND will be recovered as a special case.

Lüders' rule. Using the double-sided product as a disturbance followed by a logical AND, we find the quantum form of the *invasive conditioning rule* (13),

$$\langle \widetilde{F_X} \rangle_y = \frac{\langle y F_X y \rangle}{P(y)} = \text{Tr}(\rho_y F_X), \tag{89a}$$

$$\rho_y = \frac{y \rho y}{\text{Tr}(\rho y)}, \tag{89b}$$

for any Boolean proposition y in a framework algebra Σ_X measured prior to the observable F_X. As with the classical case, we use the tilde to indicate the intrinsic quantum invasiveness of the measurement process. If ρ and y commute, or if F_X and y commute, then the noninvasive classical conditioning rule (8) is properly recovered. This generalization of (13) is known as the projection postulate, or *Lüders' rule* [85]. If y is an atomic proposition in X, then $\rho_y = y$ as in the classical case (8) and we consistently recover the assumption (84).

For contrast, Leifer and Spekkens [67] provide a careful quantum generalization of the *noninvasive* conditioning rule (8) using a formalism based around conditional density operators. They confirm that Lüder's rule (89) cannot be obtained with pure conditioning, so it must imply additional disturbance from the measurement process itself, as indicated here.

Aharonov-Bergmann-Lebowitz rule. Just as with classical invasive conditioning, the order of conditioning will generally matter. Specifically, substituting a system proposition $z \in \Sigma_X$ into (89) yields $\langle \widetilde{z} \rangle_y = P(yzy)/P(y)$; however, $P(yzy) \neq P(zyz)$, so the "joint probability" in the numerator is order dependent unless y and z commute, just as in Eq. (26). That is, $\langle \widetilde{z} \rangle_y$ explicitly describes the case when the conditioning

proposition y is measured first as a *preselection*, followed by the proposition z.

To obtain the converse case when the conditioning proposition z is measured second as a *postselection*, we must derive the quantum form of (27). As in the classical case, we reinterpret the denominator of (89) as a marginalization $P(y) = \sum_z P(yzy)$ of the ordered joint probability that renormalizes the conditioning procedure; the identity $\sum_z z = 1_X$ permits the equality. With this interpretation, the postselected form of conditioning becomes straightforward,

$$_z\langle \widetilde{y} \rangle = \frac{P(yzy)}{\sum_{y' \in Y} P(y'zy')}. \quad (90)$$

As in the classical case, the different position of the subscript serves to distinguish the two conditioned expectations $\langle \widetilde{\cdot} \rangle_z$ and $_z\langle \widetilde{\cdot} \rangle$ corresponding to different measurement orderings.

For a pure state $\rho = x = |x\rangle\langle x|$, this postselected conditioning is known as the *Aharonov-Bergmann-Lebowitz (ABL) rule* [86], and has the form $_z\langle \widetilde{y} \rangle_x = |\langle z|y\rangle|^2 |\langle y|x\rangle|^2 / \sum_{y' \in Y} |\langle z|y'\rangle|^2 |\langle y'|x\rangle|^2$. Unlike Lüders' rule (89), the generalized ABL rule (90) does not perform a simple update to the trace-density ρ; moreover, it depends on the entire disturbance of the first measurement via the normalization sum in the denominator. If y and z commute, then the disturbance vanishes and we again correctly recover the classical case (8) that is order independent.

Bayes' Rule. The two forms of quantum invasive conditioning also lead to a modified form of Bayes' rule that relates the preselected conditioning of a sequence to the postselected conditioning of the same sequence, similarly to the classical case (29),

$$_z\langle \widetilde{y} \rangle = \langle \widetilde{z} \rangle_y \frac{P(y)}{\sum_{y' \in Y} P(y'zy')}. \quad (91)$$

If y and z commute, then the disturbance vanishes and we correctly recover Bayes' rule (10).

The unusual form of (90) has led to postselected quantum conditioning being largely overlooked. The lack of symmetry in the density update under such postselected conditioning has even prompted works in multistate-density time-symmetric reformulations of quantum mechanics [17,19–21,24,25,87], which are outside the scope of this work. However, we see here that the form of the conditioning is the same as the classically *invasive* postselected conditioning (27). Later we shall use a fully generalized form of the ABL rule (90) together with CVs to consider the subtle case of postselected averages of observables in some detail, so we delay their consideration for now.

1. Example: Polarization state

A quantum state for a single system is a classical state in some particular framework. For a two-dimensional framework such as $\{h, v\}$, all probabilities for such a classical state can be completely specified by a mixing angle θ such that $P(h) = \cos^2(\theta/2)$ and $P(v) = \sin^2(\theta/2)$. Hence, after rotating the trace-density $\rho = P(h)h + P(v)v$ to an arbitrary

framework according to (82), any quantum state trace density of polarization must have the form

$$\begin{aligned} \rho_{\theta,\beta,\gamma} &= \cos^2(\theta/2)\mathcal{U}_{\alpha,\beta,\gamma}(h) + \sin^2(\theta/2)\mathcal{U}_{\alpha,\beta,\gamma}(v) \\ &= \frac{1}{2}\begin{pmatrix} 1 + \cos\beta\cos\theta & e^{-i\gamma}\sin\beta\cos\theta \\ e^{i\gamma}\sin\beta\cos\theta & 1 - \cos\beta\cos\theta \end{pmatrix}. \quad (92) \end{aligned}$$

The α parameter of the rotation disappears in favor of the θ parameter characterizing the classical state, leaving only three net parameters, in contrast to the four parameters of an arbitrary observable (82).

The expectation functional $\langle \cdot \rangle_{\theta,\beta,\gamma}$ is then defined from the trace-density $\rho_{\theta,\beta,\gamma}$ and the unique extension of the trace Tr to the whole observable algebra $\Sigma^{\mathbb{R}}_{\mathcal{Q}(X)}$ according to $\langle F_{X'} \rangle_{\theta,\beta,\gamma} = \mathrm{Tr}(\rho_{\theta,\beta,\gamma}F_{X'})$. The trace extension is the sum of the diagonal matrix elements in the matrix representation. Hence, for the expectation of an arbitrary observable (82) under an arbitrary state (92), we find

$$\langle \mathcal{U}_{\alpha',\beta',\gamma'}(F_X) \rangle_{\theta,\beta,\gamma} = \frac{a+b}{2} + \frac{a-b}{2}(\cos\theta)\Xi, \quad (93a)$$

$$\Xi = \cos\beta\cos\beta' + \sin\beta\sin\beta'\cos(\gamma - \gamma'), \quad (93b)$$

where $\Xi \in [-1,1]$ is an interference factor that depends only on relative orientation between the state framework and the observable framework. If the frameworks coincide, then $\Xi = 1$ and the classical result is recovered.

C. Detectors and probability observables

Joint observable space. As with the classical case, we can couple a system to a detector by enlarging the sample space to the product space XY of a particular pair of frameworks. We can then perform *local* unitary rotations on each space independently to form a joint quantum sample space from the classical joint observables $\mathcal{Q}(X)\mathcal{Q}(Y)$. However, the quantum observable space also admits *global* unitary rotations on the classical joint observables to form a larger joint quantum sample space $\mathcal{Q}(XY)$. Just as with a single sample space, any two propositions in $\mathcal{Q}(XY)$ can be continuously connected with some global unitary rotation.

The full quantum observable space $\Sigma^{\mathbb{R}}_{\mathcal{Q}(XY)}$ is constructed from $\mathcal{Q}(XY)$ in the usual way. Product observables will maintain their product form under local unitary rotations, $\mathcal{U}_X(\mathcal{V}_Y(A_X B_Y)) = \mathcal{U}_X(A_X)\mathcal{V}_Y(B_Y)$. However, global unitary rotations can create unfactorable correlated joint observables in $\Sigma^{\mathbb{R}}_{\mathcal{Q}(XY)}$ even from product observables $\mathcal{U}(A_X B_Y)$.

Joint states. Similarly, joint *states* on a classical product framework extend to joint quantum states on the quantum product observable space. Under local unitary rotations, product states remain product states and classically correlated states between two specific frameworks remain classically correlated. However, *global* unitary rotations performed on any state can also form *entangled* states that have no analog in the classical theory [88]. Entangled states have some degree of *local-rotation-independent* correlation between frameworks, so display a stronger degree of correlation than can even be defined with a classically correlated state that is restricted to a single pair of frameworks. As an extreme example, maximally

entangled states are completely local-rotation independent and perfectly correlated with respect to any pair of frameworks.

Quantum operations. The specifics of entanglement do not concern us here, since any type of correlation is sufficient to represent detector probabilities within the reduced system space. For the purposes of measurement, we only assume that the correlated state with density $\rho = \mathcal{U}^{\dagger}(\rho_X \rho_Y) = U \rho_X \rho_Y U^{\dagger}$ is connected to some initial product state with density $\rho_X \rho_Y$ via a unitary rotation $\mathcal{U}^{\dagger}$. Since all quantum states can be continuously connected with some global unitary rotation that acts as a disturbance (21), this is always possible. Physically, the unitary rotation couples the known detector state ρ_Y to an unknown system state ρ_X. Furthermore, we assume that the initial state of the detector has some (not necessarily unique) pure-state expansion that is meaningful with respect to the preparation procedure $\rho_Y = \sum_{y' \in Y'} P'(y') y'$.

It then follows that the numerator for the conditioning rules (89) and (90) becomes

$$\langle y F_X y \rangle = \mathrm{Tr}(\rho y F_X y) = \mathrm{Tr}_X(\mathrm{Tr}_Y(U \rho_X \rho_Y U^{\dagger} y F_X y))$$
$$= \langle \mathcal{E}_y(F_X) \rangle_X = \mathrm{Tr}_X(\mathcal{E}_y^{\dagger}(\rho_X) F_X), \quad (94)$$

with the *operations* $\mathcal{E}_y$ and $\mathcal{E}_y^{\dagger}$ defined as

$$\mathcal{E}_y(F_X) = \langle U^{\dagger} y F_X y U \rangle_Y = \sum_{y' \in Y'} P'(y') \mathrm{Tr}_Y(y' U^{\dagger} y F_X y U)$$
$$= \sum_{y' \in Y'} M_{y,y'}^{\dagger} F_X M_{y,y'}, \quad (95a)$$

$$\mathcal{E}_y^{\dagger}(\rho_X) = \mathrm{Tr}_Y(y U \rho_X \rho_Y U^{\dagger} y)$$
$$= \sum_{y' \in Y'} P'(y') \mathrm{Tr}_Y(y U \rho_X y' U^{\dagger} y)$$
$$= \sum_{y' \in Y'} M_{y,y'} \rho_X M_{y,y'}^{\dagger}, \quad (95b)$$

$$M_{y,y'} = e^{i\phi_{y,y'}} \sqrt{P'(y')} \langle y | U | y' \rangle, \quad (95c)$$

$$M_{y,y'}^{\dagger} = e^{-i\phi_{y,y'}} \sqrt{P'(y')} \langle y' | U^{\dagger} | y \rangle. \quad (95d)$$

Here, the Hilbert space representations of the *Kraus operators* $\{M_{y,y'}\}$ have the form of partial matrix elements and are only well defined up to the arbitrary phase factors $e^{i\phi_{y,y'}}$. We also stress that $\{M_{y,y'}\}$ depend not only on the measured detector outcome y, but also on a particular detector *preparation* y'.

As a result, we find the quantum versions of the probability observables (28),

$$P(y) = \langle \mathcal{E}_y(1_X) \rangle_X = \langle E_y \rangle_X, \quad (96)$$

$$E_y = \mathcal{E}_y(1_X) = \langle U^{\dagger} y U \rangle_Y = \sum_{y' \in Y'} M_{y,y'}^{\dagger} M_{y,y'}, \quad (97)$$

and the general invasive measurement (24),

$$\langle \widetilde{F_X} \rangle_y = \frac{\langle \mathcal{E}_y(F_X) \rangle_X}{\langle \mathcal{E}_y(1_X) \rangle_X} = \frac{\sum_{y' \in Y'} \mathrm{Tr}_X(\rho_X M_{y,y'}^{\dagger} F_X M_{y,y'})}{\mathrm{Tr}_X(\rho_X E_y)}. \quad (98)$$

Similarly to the invasive classical case (25), the measurement of y on the detector must be described by a *quantum operation* $\mathcal{E}_y$ in Eq. (94), which is a completely positive map [2,6–14,16,70,89] that performs a *generalized measurement*

on the system state corresponding to the detector outcome y. The operation $\mathcal{E}_y$ acting on the identity in Eq. (97) produces a positive operator known as a *quantum effect*, E_y. By construction, the set of operations $\{\mathcal{E}_y\}$ preserves the identity, $\sum_y \mathcal{E}_y(1_X) = 1_X$; hence, the effects form a partition of the identity, $\sum_y E_y = 1_X$, making them probability observables over a particular detector framework exactly as in Eq (23).

Sequences of measurements emphasize the temporal ordering of operations, just as in the invasive classical case (26). Given two sets of quantum operations that define the sequential interaction of two detectors with the system and their subsequent conditioning, $\{\mathcal{E}_y\}$ and $\{\mathcal{E}_z'\}$, the joint probability of the ordered sequence of detector outcomes (y,z) is

$$P(y)P(z|y) = P(yzy) = P(yz1_X zy)$$
$$= \langle \mathcal{E}_y(\mathcal{E}_z'(1_X)) \rangle_X = \langle \mathcal{E}_y(E_z') \rangle_X, \quad (99)$$

where $E_z' = \mathcal{E}_z'(1_X)$. The proper sequential probability observable $\mathcal{E}_y(E_z') = \sum_{y'} M_{y,y'}^{\dagger} E_z' M_{y,y'}$ is not a simple product of the individual probability observables E_y and E_z'.

These sequence probabilities then give us the full generalization of the ABL rule (90),

$$_z\langle \widetilde{y} \rangle = \frac{\langle \mathcal{E}_y(E_z') \rangle_X}{\langle \mathcal{E}(E_z') \rangle_X} = \frac{\langle \mathcal{E}_y(E_z') \rangle_X}{\sum_{y'' \in Y} \langle \mathcal{E}_{y''}(E_z') \rangle_X}$$
$$= \frac{\sum_{y' \in Y'} \mathrm{Tr}_X(\rho_X M_{y,y'}^{\dagger} E_z' M_{y,y'})}{\sum_{y'' \in Y} \sum_{y' \in Y'} \mathrm{Tr}_X(\rho_X M_{y'',y'}^{\dagger} E_z' M_{y'',y'})}, \quad (100)$$

and the most general version of the invasive quantum Bayes' rule (91),

$$_z\langle \widetilde{y} \rangle = \langle \widetilde{E_z'} \rangle_y \frac{\langle E_y \rangle_X}{\langle \mathcal{E}(E_z') \rangle_X}. \quad (101)$$

As with (27) and (90), the postselected conditioning (100) depends on the entire disturbance of the first measurement via the *nonselective measurement* $\mathcal{E} = \sum_{y'' \in Y} \mathcal{E}_{y''}$ in the denominator.

The noncommutativity of the detection operations $\mathcal{E}_y$ emphasizes the fact that measurement is an active *process*: an experimenter alters the quantum state by coupling it to a detector and then conditioning on acquired information from the detector. Without some filtering process that completes the disturbance implied by (94), there is no measurement. The nonselective measurement $\mathcal{E}$ also includes the active disturbance of the measurement process, but does not condition on a particular outcome. Furthermore, measuring a quantum state in a different order generally disturbs it differently. The state may also, in certain conditions, be probabilistically "uncollapsed" back to where it started by using the correct conditioning sequence [3–5]. In this sense, sequential quantum conditioning is analogous to a stochastic control process that guides the progressive disturbance of a state along some trajectory in the state space [16].

Measurement operators. Since the quantum operation $\mathcal{E}_y$ performs a measurement, we will refer to its Kraus operators $\{M_{y,y'}\}$ (95) as *measurement operators*. However, a quantum operation generally has many equivalent double-sided product expansions like (95a) in terms of measurement operators. Each such set of measurement operators $\{M_{y,y'}\}$ corresponds to a

specific choice of framework for the preparation of the detector state $\rho_Y = \sum_{y' \in Y'} P(y')y'$.

Given a specific set of measurement operators, the substitution $M_{y,y'} \rightarrow U_{y,y'}M_{y,y'}$ with unitary $U_{y,y'}$ will produce the same effect E_y according to (97) but will correspond to a different operation $\mathcal{E}'_y$. Hence, we conclude that many measurement operations can produce the same probability observables on the system space [90]. Therefore, probability observables are not sufficient to completely specify a quantum measurement: one needs to specify the full operations as in the classically invasive case (25).

Quantum process tomography. Just as classical probability observables can be characterized via process tomography, operations can be characterized by *quantum process tomography.* One performs quantum process tomography by sending known states into a detector, measuring the detector, then measuring the resulting states to see how the state was changed by the detector. Since quantum operations contain information about disturbance as well as conditioning, quantum process tomography generally requires more characterization measurements than pure classical process tomography.

Pure operations. An initially pure detector state with density y' produces a *pure operation* $\mathcal{E}_y(F_X) = M_y^\dagger F_X M_y$ with a single associated measurement operator $M_y = e^{i\phi_y}\langle y|U|y'\rangle$ that is unique up to the arbitrary phase factor $e^{i\phi_y}$. Most laboratory preparation procedures for the detector are designed to produce a pure initial state, so pure operations will be the typical case. A pure operation has the additional property of partially collapsing a pure state to another pure state. It is also most directly related to the probability observable $E_y = M_y^\dagger M_y$, since the single-measurement operator has a polar decomposition $M_y = U_y E_y^{1/2}$ in terms of the positive root of the probability observable $E_y^{1/2}$.

Weak measurement. If we wish for such a conditioning process to leave the state approximately unchanged, we must make a *weak measurement*, just as in the classical case (19). However, a quantum weak measurement requires a strict condition regarding the measurement operations and not just the probability observables due to the additional disturbance in the measurement. Formally, the measurement operations typically depend on a measurement strength parameter ϵ such that

$$\forall\, y \in Y, \;\; \lim_{\epsilon \to 0} \mathcal{E}_y(\epsilon; F_X) = P_Y(y)\mathcal{I}(F_X), \quad (102)$$

where $\mathcal{I}$ is the identity operation and $P_Y(y)$ is the probability for obtaining the detector outcome y in the absence of interaction. As with the classical case, the limit as $\epsilon \to 0$ is an idealization known as the *weak-measurement limit* and is not strictly achievable in the laboratory.

The definition (102) implies that subsequent measurements will be unaffected, $\forall\, y \in Y$, $\lim_{\epsilon \to 0}\langle \widetilde{F_X}\rangle_y = \langle F_X\rangle$, and that the probability observables are proportional to the identity in the weak limit, $\forall\, y \in Y$, $\lim_{\epsilon \to 0} E_y(\epsilon) = P_Y(y)1_X$, just as in the classical case (19). It also follows that any set of measurement operators $\{M_{y,y'}(\epsilon)\}$ that characterize $\mathcal{E}_y(\epsilon)$ must also be proportional to the identity in the weak limit $\forall\, y \in Y$, $y' \in Y'$, $\lim_{\epsilon \to 0} M_{y,y'}(\epsilon) \propto 1_X$.

Weak measurements are more interesting in the quantum case than in the classical case due to the existence of incompat-ible frameworks. Since a weak measurement of an observable does not appreciably affect the quantum state, subsequent measurements on incompatible observables can be made that will probe approximately the same state. This technique allows (noisy) information about two incompatible frameworks to be gleaned from nearly the same quantum state in a single experiment, which is strictly impossible using strong measurements that collapse the state to a pure state in a particular framework after each measurement. The penalty for using weak measurements is that many more measurements are needed than in the strong-measurement case to overcome the ambiguity of the measurement, as discussed in the classical case.

1. Example: Coverslip polarization detector

To cement these ideas, we consider the task of indirectly measuring polarization in a particular framework. For specificity, we will consider the passage of a laser beam with unknown polarization through a glass microscope coverslip, as shown in Fig. 2. Fresnel reflection off the coverslip leads to a disparity between transmission and reflection of the polarizations, so comparing transmitted to reflected light allows a generalized measurement of polarization, as we demonstrated experimentally in Ref. [51].

The system sample space we wish to measure is the polarization with respect to the table ($h = |h\rangle\langle h|$) and ($v = |v\rangle\langle v|$), which could in principle be measured ideally with a polarizing beam splitter. The detector sample space is the spatial degree of freedom of the transmitted ($t = |t\rangle\langle t|$) and reflected ($r = |r\rangle\langle r|$) ports of a coverslip rotated to some fixed angle with respect to the incident beam around an axis perpendicular to the table. The initial state of the detector is the pure state indicating that the beam enters a single incident port ($b = |b\rangle\langle b|$) of the coverslip with certainty. The rotation $\mathcal{U}^\dagger(\rho_X b) = U\rho_X b U^\dagger$ that couples the system to the detector describes the interaction of the beam with the coverslip and has a unitary rotor U corresponding to the polarization-dependent scattering matrix of the coverslip. Assuming that the scattering preserves beams of pure polarization, so h remains h and v
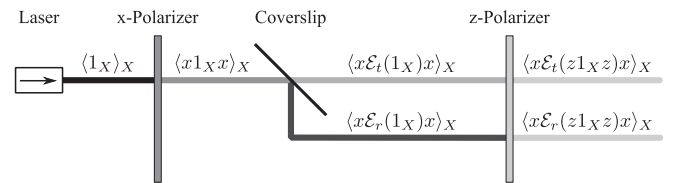

FIG. 2. Coverslip polarization measurement. A laser beam passes through a preselection x polarizer, a glass microscope coverslip, and a postselection z polarizer. The transmission probabilities for each segment of the apparatus are shown. By assigning appropriate contextual values $f_Y(t)$ and $f_Y(r)$ (12) to the output ports of the coverslip, the polarization observable $F_X = f_X(h)h + f_X(v)v$ can be measured using the equivalent measurement context $F_X = f_Y(t)\mathcal{E}_t(1_X) + f_Y(r)\mathcal{E}_r(1_X)$. Averaging the same contextual values with pre- and postselected conditional probabilities $_z\langle \widetilde{t}\rangle_x = \langle x\mathcal{E}_t(z)x\rangle_X/[\langle x\mathcal{E}_t(z)x\rangle_X + \langle x\mathcal{E}_r(z)x\rangle_X]$ and $_z\langle \widetilde{r}\rangle_x = \langle x\mathcal{E}_r(z)x\rangle_X/[\langle x\mathcal{E}_t(z)x\rangle_X + \langle x\mathcal{E}_r(z)x\rangle_X]$ produces the conditioned average (127) $_z\langle \widetilde{F_X}\rangle_x = f_Y(t)_z\langle \widetilde{t}\rangle_x + f_Y(r)_z\langle \widetilde{r}\rangle_x$.

remains v, the rotor decouples into a direct sum of rotors that are specific to each polarization,

$$U = U_h \oplus U_v, \qquad (103)$$

meaning that U has a block-diagonal structure when represented as a matrix.

Selecting each output port of the coverslip produces the two measurement operators according to (95),

$$M_t = \langle t|U|b\rangle = \begin{pmatrix} \langle t|U_h|b\rangle & 0 \\ 0 & \langle t|U_v|b\rangle \end{pmatrix}, \qquad (104a)$$

$$M_r = \langle r|U|b\rangle = \begin{pmatrix} \langle r|U_h|b\rangle & 0 \\ 0 & \langle r|U_v|b\rangle \end{pmatrix}, \qquad (104b)$$

which characterize the *pure measurement operations* that modify observables according to (95a)

$$\mathcal{E}_t(F_X) = M_t^\dagger F_X M_t, \qquad (105a)$$

$$\mathcal{E}_r(F_X) = M_r^\dagger F_X M_r, \qquad (105b)$$

and their adjoints that modify the state density according to (95b)

$$\mathcal{E}_t^\dagger(\rho_X) = M_t \rho_X M_t^\dagger, \qquad (106a)$$

$$\mathcal{E}_r^\dagger(\rho_X) = M_r \rho_X M_r^\dagger. \qquad (106b)$$

The pure measurement operations in turn produce *probability observables* according to (97)

$$E_t = \mathcal{E}_t(1_X) = M_t^\dagger M_t = \begin{pmatrix} |\langle t|U_h|b\rangle|^2 & 0 \\ 0 & |\langle t|U_v|b\rangle|^2 \end{pmatrix}, \qquad (107a)$$

$$E_r = \mathcal{E}_r(1_X) = M_r^\dagger M_r = \begin{pmatrix} |\langle r|U_h|b\rangle|^2 & 0 \\ 0 & |\langle r|U_v|b\rangle|^2 \end{pmatrix}, \qquad (107b)$$

in the same framework as h and v. These probability observables are therefore equivalent to classical probability observables (23) specified by the effective characterization probabilities $\tilde{P}(t|h) = |\langle t|U_h|b\rangle|^2$, $\tilde{P}(r|h) = |\langle r|U_h|b\rangle|^2$, $\tilde{P}(t|v) = |\langle t|U_v|b\rangle|^2$, and $\tilde{P}(r|v) = |\langle r|U_v|b\rangle|^2$.

The measurement operators (104) have a polar decomposition in terms of the roots of the probability observables and an extra unitary phase contribution,

$$M_t = \begin{pmatrix} e^{i\phi_{h,t}}\sqrt{\tilde{P}(t|h)} & 0 \\ 0 & e^{i\phi_{v,t}}\sqrt{\tilde{P}(t|v)} \end{pmatrix}, \qquad (108a)$$

$$M_r = \begin{pmatrix} e^{i\phi_{h,r}}\sqrt{\tilde{P}(r|h)} & 0 \\ 0 & e^{i\phi_{v,r}}\sqrt{\tilde{P}(r|v)} \end{pmatrix}. \qquad (108b)$$

Any nonzero relative phase, such as $\phi_{h,t} - \phi_{v,t}$, will affect the framework orientation for subsequent measurements; however, it will not contribute to the acquisition of information from the measurement since it does not contribute to the probability observables. Such relative phase is therefore part of the *disturbance* of the measurement process.

Specifically, the initial state of polarization P_X will be conditioned by a selection of a particular port on the detector according to

$$\langle \widetilde{F_X} \rangle_t = \frac{\langle \mathcal{E}_t(F_X) \rangle_X}{\langle \mathcal{E}_t(1_X) \rangle_X} = \frac{\mathrm{Tr}_X(M_t \rho_X M_t^\dagger F_X)}{\mathrm{Tr}_X(\rho_X E_t)}, \qquad (109a)$$

$$\langle \widetilde{F_X} \rangle_r = \frac{\langle \mathcal{E}_r(F_X) \rangle_X}{\langle \mathcal{E}_r(1_X) \rangle_X} = \frac{\mathrm{Tr}_X(M_r \rho_X M_r^\dagger F_X)}{\mathrm{Tr}_X(\rho_X E_r)}. \qquad (109b)$$

Although the probabilities in each denominator only depend on the probability observables, the altered states in each numerator depend on the measurement operations and will include effects from the relative phase in the measurement operators (108).

D. Contextual values

Operation correspondence. The introduction of contextual values in the quantum case proceeds identically to the classical case of invasive measurements (34). Since we must generally represent detector probabilities by *operations* $\{\mathcal{E}_y\}$ within the reduced system space according to (97) and (99), we must also generally represent detector observables by *weighted operations* within the reduced system space,

$$\langle F_Y \rangle = \sum_{y \in Y} f_Y(y)P(y) = \sum_{y \in Y} f_Y(y)\langle \mathcal{E}_y(1_X)\rangle_X = \langle \mathcal{F}_X(1_X)\rangle_X, \qquad (110)$$

$$\mathcal{F}_X = \sum_{y \in Y} f_Y(y)\mathcal{E}_y. \qquad (111)$$

If we are concerned with only a single measurement or are working within a single framework as in the classical formalism then, for all practical purposes, the operation $\mathcal{F}_X$ reduces to its associated system observable $F_X = \mathcal{F}_X(1_X)$ as in the classical definition (31).

Contextual values. We observe a corollary exactly as in the classical case (32): If we can expand a *system* observable in terms of the probability observables generated by a particular measurement operation, then that observable can also be expressed as an equivalent *detector* observable,

$$F_X = \sum_{y \in Y} f_Y(y)E_y \Rightarrow F_Y = \sum_y f_Y(y)y, \qquad (112)$$

which is the quantum form of our main result originally introduced in Ref. [49]. As in the classical case, we dub the required detector labels $f_Y(y)$ the contextual values (CVs) of the quantum observable F_X with respect to the *context* of a specific detection scheme as represented in the system space by the measurement operations $\{\mathcal{E}_y\}$. Since many measurement operations produce the same probability observables $\{\mathcal{E}_y(1_X) = E_y\}$, many detection schemes can use the same CVs to reproduce an observable average.

Moments. As with classically invasive measurements (35), higher statistical moments of the observable require more care to measure. For instance, we require the following equality in order to accurately reproduce the nth moment of an observable indirectly using the same CVs:

$$\langle (F_X)^n \rangle_X = \sum_{y_1,\dots,y_n \in Y} f_Y(y_1)\cdots f_Y(y_n)\langle E_{y_1}\cdots E_{y_n}\rangle_X. \qquad (113)$$

However, as indicated in Eq. (99), performing a sequence of n measurements produces the measurable probability $\langle \mathcal{E}_{y_1}(\cdots(E_{y_n})\cdots)\rangle_X \neq \langle E_{y_1}\cdots E_{y_n}\rangle_X$. Indeed, $\langle E_{y_1}\cdots E_{y_n}\rangle_X$ will not generally be a well-formed probability. To obtain the equality (113) with a particular choice of CVs, we need the additional constraint that all the measurement operators must commute with each other. As a result, they must be part of the same framework as the system observable and hence commute with that observable as well. We will call any detector with commuting measurement operators with respect to a particular observable a *fully compatible detector* for that observable. Evidently, this is a strict requirement for a detector.

Alternatively, as with the classical case, we can change the CVs to define new observables that correspond to powers of the original observable, such as $G_X = (F_X)^n = \sum_{y \in Y} g_Y(y) E_y$. These new observables can then be measured indirectly using the same experimental setup without the need for measurement sequences. The CVs $g_Y(y)$ for the nth power of F_X will not be a simple power of the CVs $f_Y(y)$ for F_X unless the measurement is unambiguous.

Correlation functions. If a time-evolution unitary rotation $\mathcal{U}_t$ is inserted between different observable measurements, then we obtain a quantum *correlation function* instead,

$$\langle \widetilde{F_X(0)G_X}(t)\rangle = \langle \mathcal{F}_X(\mathcal{U}_t(\mathcal{G}_X(1_X)))\rangle_X , \qquad (114)$$

which should be compared to the classical case (36). Similarly, n-time correlations can be defined with $n - 1$ time-evolutions between the observable measurements $\langle \mathcal{F}_1(\mathcal{U}_{t_1}(\mathcal{F}_2(\cdots \mathcal{U}_{t_{n-1}}(\mathcal{F}_n(1_X))\cdots)))\rangle$.

Inversion. Since the CVs depend only on the probability observables, which commute with the measured observable for a fully compatible detector, the procedure for determining the CVs will be identical to the classical case. That is, the contextual values of a quantum observable exactly correspond to the detector labels for a classically ambiguous detector. We shall refer the reader back to the classical inversion (40) for discussion on how to solve the relation (112). As a reminder, we advocate the pseudoinverse as a principled approach for picking the CVs in the event of redundancy or course-graining.

Conditioned averages. We can construct a general *postselected conditioned average* from the CVs and the fully generalized ABL rule (100) analogously to the classical case (37),

$$_z\langle \widetilde{F_X}\rangle = \sum_y f_Y(y)_z\langle \tilde{y}\rangle = \frac{\langle \mathcal{F}_X(E_z')\rangle_X}{\langle \mathcal{E}(E_z')\rangle_X}$$

$$= \frac{\sum_{y \in Y} \sum_{y' \in Y'} f_Y(y)\mathrm{Tr}(\rho_X M_{y,y'}^\dagger E_z' M_{y,y'})}{\sum_{y \in Y} \sum_{y' \in Y'} \mathrm{Tr}(\rho_X M_{y,y'}^\dagger E_z' M_{y,y'})}. \quad (115)$$

We introduced this type of conditioned average in Ref. [49] for the typical case of pure operations $\{\mathcal{E}_y\}$ with single associated measurement operators $\{M_y\}$.

If the postselection is defined in the same framework as the measurement operation, then the nonselective measurement $\mathcal{E}$ in the denominator will reduce to unity, leaving a classical conditioned average

$$\langle F_X\rangle_z = \frac{\sum_{y \in Y} f_Y(y)\langle E_y E_z'\rangle_X}{\langle E_z'\rangle_X} = \frac{\langle F_X E_z'\rangle_X}{\langle E_z'\rangle_X}, \quad (116)$$

of the same form as (18). Similarly, the preselected conditioning (98) will also reduce to (116) for such a case. This special case cannot exceed the eigenvalue range of the observable: the observable F_X will always reduce to its eigenvalues since either the state or the postselection commute with it.

More generally, however, the combination of amplified CVs and the context-dependent probabilities in the general postselected average (115) can send it outside the eigenvalue range of the observable. As we discussed in Refs. [51,91], having such a conditioned average stray outside the eigenvalue range of the observable is equivalent to a violation of a Leggett-Garg inequality that tests the assumptions of macrorealism under noninvasive detection. As a result, an eigenvalue range violation gives a direct indication of either *nonclassicality* present in a measurement sequence, or intrinsic measurement *disturbance* beyond that of noninvasive classical conditioning as we saw in the example in Sec. II D 2. We refer the reader to [51,91] for more detail on this matter.

Strong-conditioned average. There are two other important special cases of the conditioned average (115) worth mentioning: strong measurement and weak measurement. The strong-measurement case is distinguished by being constrained exclusively to the eigenvalue range of the observable. Specifically, (115) reduces to the form

$$_z\langle \widetilde{F_X}\rangle = \frac{\sum_{x \in X} f_X(x)P(x)D_x(z)}{\sum_{x \in X} P(x)D_x(z)}$$

$$= \frac{\sum_{x \in X} f_X(x)\langle x|\rho|x\rangle|\langle x|z\rangle|^2}{\sum_{x \in X} \langle x|\rho|x\rangle|\langle x|z\rangle|^2}, \quad (117)$$

which contains only the eigenvalues $f_X(x)$ of the observable and factored probability products. However, it cannot be expressed solely in terms of the observable F_X and a conditioned state as in the classical case (37) due to the disturbances $D_x(z)$. Only when the state or postselection commutes with the observable does (117) reduce to a special case of (116) and become free from disturbance.

Weak values. The weak-measurement case is distinguished by being the only case of the quantum postselected conditioned average (115) that can become *context independent* for any state and postselection (under certain conditions). The context-independent weak limit of the conditioned average (115) is the *weak value* [17,19–21,24,25,49],

$$_z\langle \widetilde{F_X}\rangle^w = \frac{\langle E_z' F_X + F_X E_z'\rangle_X}{2\langle E_z'\rangle_X}, \quad (118)$$

and is expressed entirely in terms of the system expectation functional $\langle \cdot \rangle_X$, the postselection probability observable E_z', and the observable F_X. Written in this form it is clear that it is a symmetrized version of the context-independent commuting case (116); however, unlike (116) the weak value, (118) is not constrained to the eigenvalue range and can even diverge. For a pure initial state with trace-density x and pure postselection z, the weak value (118) takes the traditional form

$$_z\langle \widetilde{F_X}\rangle_x^w \to \mathrm{Re}\frac{\langle z|F_X|x\rangle}{\langle z|x\rangle}. \quad (119)$$

We will consider under what conditions one can obtain such a weak value in Sec. III E.

1. Example: Coverslip detector revisited

Continuing the example from Sec. III C 1 and Fig. 2, observables defined in the same framework as the probability observables may be expressed in terms of the probability observables according to (112) using *contextual values* (CVs), exactly as in the classical example (47),

$$F_X = f_X(h)h + f_X(v)v = f_Y(t)E_t + f_Y(r)E_r, \quad (120a)$$

$$\begin{pmatrix} f_X(h) \\ f_X(v) \end{pmatrix} = \begin{pmatrix} \tilde{P}(t|h) & \tilde{P}(r|h) \\ \tilde{P}(t|v) & \tilde{P}(r|v) \end{pmatrix} \begin{pmatrix} f_Y(t) \\ f_Y(r) \end{pmatrix}. \quad (120b)$$

Inverting this relation according to (40) produces the unique CV,

$$f_Y(t) = \frac{\tilde{P}(r|v)f_X(h) - \tilde{P}(r|h)f_X(v)}{\tilde{P}(t|h)\tilde{P}(r|v) - \tilde{P}(r|h)\tilde{P}(t|v)}, \quad (121a)$$

$$f_Y(r) = -\frac{\tilde{P}(t|v)f_X(h) - \tilde{P}(t|h)f_X(v)}{\tilde{P}(t|h)\tilde{P}(r|v) - \tilde{P}(r|h)\tilde{P}(t|v)}. \quad (121b)$$

The denominator is unity when the output ports of the coverslip are perfectly correlated with the polarization. Otherwise, the denominator is less than one and serves to *amplify* the CV to compensate for the ambiguity of the detection. The numerator contains cross-compensation factors that correct bias in the detector; that is, the eigenvalue $f_X(h)$ for h in the contextual value $f_Y(t)$ for t is weighted by the conditional probability $\tilde{P}(r|v)$ corresponding to the complementary quantities of v and r, and so forth.

The CVs define the detector observable that is actually being measured in the laboratory,

$$F_Y = f_Y(t)t + f_Y(r)r. \quad (122)$$

This detector observable corresponds to a detection *operation* on the system space according to (111)

$$\mathcal{F}_X = f_Y(t)\mathcal{E}_t + f_Y(r)\mathcal{E}_r, \quad (123)$$

which fully describes the interaction with the detector, subsequent conditioning, and experimental convention for defining the observable. When no subsequent conditioning is performed on the system, this operation constructs the system observable $F_X = \mathcal{F}_X(1_X) = f_Y(t)E_t + f_Y(r)E_r$, as desired.

Since the pure measurement operations all belong to the same framework and commute with F_X, the operation $\mathcal{F}_X$ is also *fully compatible* with the observable F_X, meaning it can measure any moment of that observable using the same CV according to (113),

$$\langle \mathcal{F}_X^n(1_X) \rangle_X = \langle (F_X)^n \rangle_X = \sum_{i_1 \dots i_n} f_Y(i_1) \cdots f_Y(i_n) \langle E_{i_1} \cdots E_{i_n} \rangle_X. \quad (124)$$

The quantity $\mathcal{F}_X^n(1_X)$ indicates a sequence of n consecutive measurements made by the same coverslip on the beam to construct the observable $(F_X)^n$ for the nth moment of F_X. That is, the output from each port of the coverslip is fed back into the coverslip to be measured again. There are 2^n possible outcome sequences $(i_1, \dots, i_n)$ for n traversals through the coverslip, each with probability $\langle E_{i_1} \cdots E_{i_n} \rangle_X$ of occurring. These probabilities are weighted with appropriate products of corresponding CVs and summed to correctly construct the nth moment of F_X.

Alternatively, one can change the CV to directly measure the observable $G_X = (F_X)^n = g_Y(t)E_t + g_Y(r)E_r$ from one traversal of the coverslip. The required CVs for G_X,

$$g_Y(t) = \frac{\tilde{P}(r|v)[f_X(h)]^n - \tilde{P}(r|h)[f_X(v)]^n}{\tilde{P}(t|h)\tilde{P}(r|v) - \tilde{P}(r|h)\tilde{P}(t|v)}, \quad (125a)$$

$$g_Y(r) = -\frac{\tilde{P}(t|v)[f_X(h)]^n - \tilde{P}(t|h)[f_X(v)]^n}{\tilde{P}(t|h)\tilde{P}(r|v) - \tilde{P}(r|h)\tilde{P}(t|v)}, \quad (125b)$$

are not simple powers of the CV (121) for F_X unless the measurement is unambiguous.

In addition to moments of F_X, we can obtain postselected *conditioned averages* of F_X by conditioning on a second measurement outcome characterized by a probability observable E_z' after the measurement by the coverslip according to (115),

$$_z\langle \widetilde{F_X} \rangle = \frac{\langle \mathcal{F}_X(E_z') \rangle_X}{\langle \mathcal{E}(E_z') \rangle_X}, \quad (126)$$

where $\mathcal{E} = \mathcal{E}_t + \mathcal{E}_r$ is the nonselective measurement by the coverslip. The second measurement could be a polarizer, another coverslip, or any other method for measuring polarization a second time.

If the initial state is pure with a density $\rho = x = |x\rangle\langle x|$ and the final postselection is also pure $z = |z\rangle\langle z|$, then (126) simplifies to a pre- and postselected conditioned average,

$$_z\langle \widetilde{F_X} \rangle_x = \frac{f_Y(t)|\langle z|M_t|x\rangle|^2 + f_Y(r)|\langle z|M_r|x\rangle|^2}{|\langle z|M_t|x\rangle|^2 + |\langle z|M_r|x\rangle|^2}. \quad (127)$$

If we relate both pure states to the reference state h via unitary rotations as defined in Eq. (79), $x = \mathcal{U}_{\alpha,\beta,\gamma}(h)$ and $z = \mathcal{U}_{\alpha',\beta',\gamma'}(h)$, then the probabilities take the form

$$\begin{aligned} |\langle z|M_t|x\rangle|^2 &= \tilde{P}^h(t)\cos^2(\beta/2)\cos^2(\beta'/2) \quad (128a) \\ &+ \tilde{P}^v(t)\sin^2(\beta/2)\sin^2(\beta'/2) \\ &+ \frac{\sqrt{\tilde{P}^h(t)\tilde{P}^v(t)}}{2}\sin\beta\sin\beta' \\ &\times \cos(\gamma - \gamma' - \phi_{h,t} + \phi_{v,t}), \\ |\langle z|M_r|x\rangle|^2 &= \tilde{P}^h(r)\cos^2(\beta/2)\cos^2(\beta'/2) \quad (128b) \\ &+ \tilde{P}^v(r)\sin^2(\beta/2)\sin^2(\beta'/2) \\ &+ \frac{\sqrt{\tilde{P}^h(r)\tilde{P}^v(r)}}{2}\sin\beta\sin\beta' \\ &\times \cos(\gamma - \gamma' - \phi_{h,r} + \phi_{v,r}). \end{aligned}$$

We see that each probability possesses an interference term that stems from the relative orientations of the incompatible frameworks for the preparation, measurement, and postselection. In addition, the relative phases in the measurement operators (108) will affect the orientations of the frameworks and further disturb the measurement, as mentioned. For the classical case, the frameworks coincide, so $\beta,\beta' \in \{0,\pi\}$; the interference term vanishes and the probabilities reduce to the conditional probabilities that characterize the probability observables.

The combination of the expanded range of the CVs (121) and the interference term in the probabilities (128) can make the postselected conditioned averages (126) counterintuitively exceed the eigenvalue range of the observable F_X. Such a

violation of the eigenvalue range cannot occur from classical conditioning without disturbance as in Sec. II D 2.

2. Example: Calcite polarization detector

We can also measure polarization using a von Neumann measurement [2] that uses a detector with a continuous sample space detector, such as position. For example, passing a beam of polarized light through a calcite crystal will continuously separate the polarizations h and v along a particular position axis. Measuring the position profile of the resulting split beam along that axis allows information to be gained about the polarization.

For such a setup, measuring the position with a linear scale corresponds to measuring a detector observable $Q = \int_Y y d|y\rangle\langle y|$ for a continuous sample space of distinguishable positions. The observable Q has a conjugate D_Q that satisfies $[Q, D_Q] = i1_Y$. The conjugate can thus generate translations in Q with a unitary rotor, $\exp(iqD_Q)Q\exp(-iqD_Q) = Q + [iqD_Q, Q] + [iqD_Q, [iqD_Q, Q]] + \cdots = Q + q1_Y$. Hence, we can model the calcite crystal as a rotation governed by a unitary rotor of the form

$$U = \exp(-i(\epsilon_h h - \epsilon_v v)D_Q), \quad (129)$$

which will translate h polarization by some amount ϵ_h while simultaneously translating v polarization by some amount ϵ_v in the opposing direction. The parameters ϵ_h and ϵ_v will depend on the geometry of the crystal with respect to the incident beam.

Suppose the light beam has an initially pure beam profile state described by a density $\rho = |\psi\rangle\langle\psi|$. The probability for obtaining a particular pure position $y = |y\rangle\langle y|$ in the profile would then be $dP_Y(y) = p_Y(y)dy = \mathrm{Tr}(\rho y)dy = |\langle y|\psi\rangle|^2 dy$. Each complex factor $\langle y|\psi\rangle$ is the "wave function" of the transverse beam profile, whose complex square is the probability density with respect to the integral $p_Y(y) = |\langle y|\psi\rangle|^2$.

If we then pass the beam through the crystal described by the rotor (129) and measure its position in a pure position state $y = |y\rangle\langle y|$, we will have enacted a pure operation on the polarization of the beam that is characterized by a single measurement operator,

$$d\mathcal{E}_y(F_X) = M(y)^\dagger F_X M(y)dy, \quad (130a)$$

$$M(y) = \langle y|U|\psi\rangle = h\langle y - \epsilon_h|\psi\rangle + v\langle y + \epsilon_v|\psi\rangle, \quad (130b)$$

with components equal to the initial wave function of the detector profile shifted in position by an appropriate ϵ. The pure measurement operations define a continuous set of probability observables,

$$dE(y) = d\mathcal{E}_y(1_X) = M(y)^\dagger M(y)dy$$
$$= h dP_Y(y - \epsilon_h) + v dP_Y(y + \epsilon_v), \quad (131)$$

with components equal to the initial transverse beam profile shifted in position by an appropriate ϵ. Unless the shifts become degenerate with $\epsilon_v = -\epsilon_h$ then these probability observables can be used to indirectly measure any observable in the framework of h and v.

Since the observable $\epsilon_h h - \epsilon_v v$ appears as a generator for the rotation U, it could be tempting to assert that the detector must specifically measure this observable. However, only the *framework* in which the generating observable is defined determines which observables can be measured. The choice of CVs, which can be made in postprocessing, will calibrate the detector to measure specific observables in that framework.

We considered a classical version of similar probability observables in Sec. II D 4. Generalizing that derivation only slightly, we can find the preferred contextual values (CVs) $f_Y(y)$ for an arbitrary polarization observable $F_X = f_X(h)h + f_X(v)v$,

$$f_Y(y) = f_X(h)\frac{v_+(y) + v_-(y)}{2} + f_X(v)\frac{v_+(y) - v_-(y)}{2}, \quad (132a)$$

$$v_+(y) = \frac{p_Y(y - \epsilon_h) + p_Y(y + \epsilon_v)}{a + b(\epsilon_h, \epsilon_v)}, \quad (132b)$$

$$v_-(y) = \frac{p_Y(y - \epsilon_h) - p_Y(y + \epsilon_v)}{a - b(\epsilon_h, \epsilon_v)}, \quad (132c)$$

$$a = \int_Y p_Y^2(y)dy, \quad (132d)$$

$$b(\epsilon_h, \epsilon_v) = \int_Y p_Y(y - \epsilon_h)p_Y(y + \epsilon_v)dy. \quad (132e)$$

In particular, one can measure the orthogonal observables $h - v$ and 1_X using the expansions

$$h - v = \int_Y v_-(y)dE(y), \quad (133)$$

$$1_X = h + v = \int_Y v_+(q)dE(y). \quad (134)$$

For the specific case of an initial Gaussian beam centered at zero, we have

$$p(y) = \exp\left(-\frac{y^2}{2\sigma^2}\right)\Big/\sigma\sqrt{2\pi}, \quad (135a)$$

$$\epsilon = (\epsilon_h + \epsilon_v)/2, \quad (135b)$$

$$\delta = (\epsilon_h - \epsilon_v)/2, \quad (135c)$$

$$a = \frac{1}{2\sigma\sqrt{\pi}}, \quad (135d)$$

$$b(\epsilon) = a\exp(-(\epsilon/\sigma)^2), \quad (135e)$$

$$v_-(y) = \sqrt{2}\frac{\exp\left(-\frac{(y-\delta)^2}{2\sigma^2}\right)\sinh\left(\frac{\epsilon(y-\delta)}{\sigma^2}\right)}{\sinh\left(\frac{\epsilon^2}{2\sigma^2}\right)}, \quad (135f)$$

$$v_+(y) = \sqrt{2}\frac{\exp\left(-\frac{(y-\delta)^2}{2\sigma^2}\right)\cosh\left(\frac{\epsilon(y-\delta)}{\sigma^2}\right)}{\cosh\left(\frac{\epsilon^2}{2\sigma^2}\right)}. \quad (135g)$$

What matters for the measurement is the average translation ϵ away from the midpoint $(y - \delta)$. The amplification of the CVs is controlled by the parameter ϵ/σ, which serves as an indicator for the ambiguity of the measurement. When the shift ϵ is large compared to the width of the Gaussian σ, then $\epsilon/\sigma \gg 1$; the shifted Gaussians for h and v are distinguishable; the CVs approach the eigenvalues of the measurement; and, the measurement is unambiguous. When the shift is small compared to the width of the Gaussian, then $\epsilon/\sigma \ll 1$, the
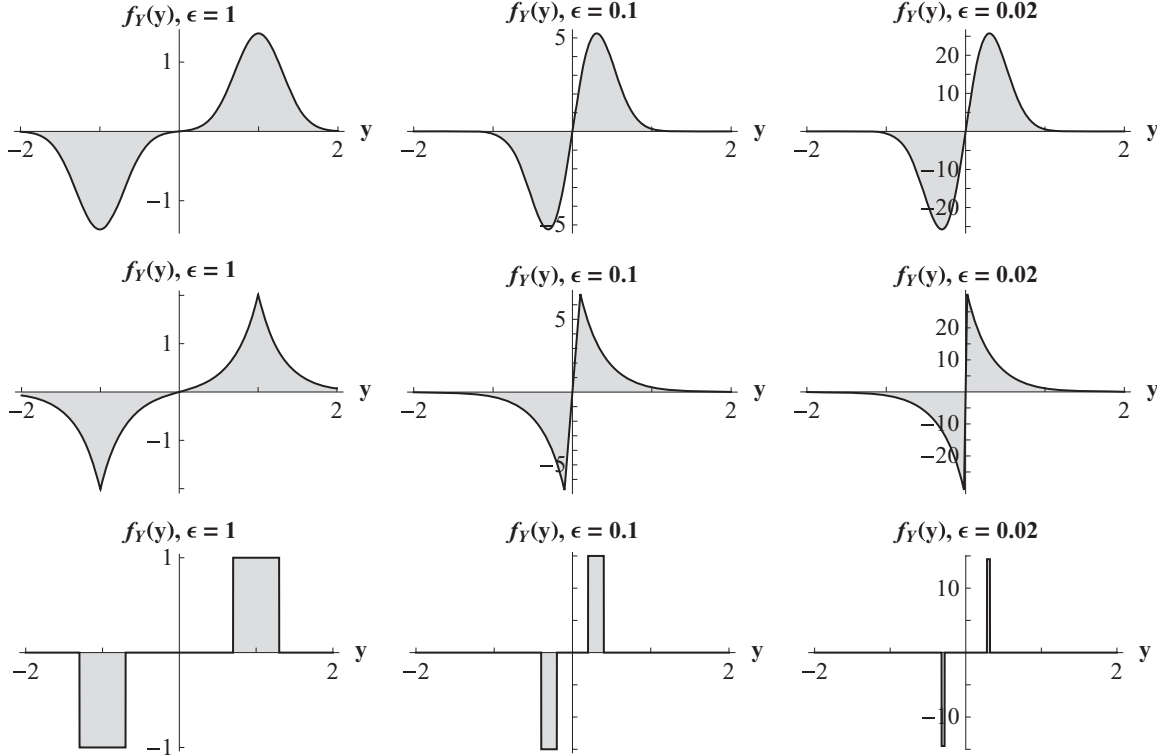
FIG. 3. Preferred CVs $f_Y(y)$ given in Eq. (132c) for a calcite position measurement that targets the polarization observable $F_X = h - v$, shown for strong separation ($\epsilon = 1$), wimpy separation ($\epsilon = 0.1$), and weak separation ($\epsilon = 0.02$) of the polarizations. (top row) Initial Gaussian beam profile. (middle row) Initial Laplace beam profile. (bottom row) Initial top-hat beam profile. Note that the top-hat CVs are the eigenvalues of ± 1 under strong separation, but become amplified as the distributions start to overlap; moreover, the top-hat CVs cancel out in the perfectly ambiguous overlapping region. The amplification and cancellation behavior of the CVs is more complicated for less-definite detector profiles.

Gaussians for h and v largely overlap, the CVs diverge, and the measurement is ambiguous. Figure 3 shows the CVs (135f) for the Gaussian initial beam profile, as well as for a Laplace and top-hat profile for comparison.

This sort of detection protocol was used in the original paper on weak values [17] in the form of a Stern-Gerlach apparatus that measures spin analogously to polarization using a continuous momentum displacement generated by a magnetic field. The initial Gaussian beam profile shifted an amount ϵ away from the midpoint of the initial beam profile in a direction corresponding to the value of the spin. Since the beam profile was symmetric about its mean, the generic CVs $f_Y(y) = y/\epsilon$ were implicitly assigned as a linear calibration of the detector, which targets a specific observable analogous to $h - v$. Motivating this implicit choice was the fact that, when ϵ is sufficiently small, the two overlapping Gaussians produce, to a good approximation, a single resulting Gaussian with a shifted mean consistent with such a linear scaling, as shown in Fig. 4. That such a choice was being made was later pointed out explicitly in Ref. [23] before we identified the role of the CVs in Ref. [49] and derived the preferred form (135f). The proposed spin-measurement protocol was adapted to a polarization measurement using a calcite crystal, as we have developed in this section, and then verified experimentally [18,34].

To produce the weak value from the polarization measurement, we postselect on a second measurement to form a conditioned average. If the initial polarization state is pure

with a density $\rho = x = |x\rangle\langle x|$ and the final postselection is also pure $z = |z\rangle\langle z|$, then we have the form

$$_z\langle\widetilde{F_X}\rangle_x = \frac{\int_Y f_Y(y)|\langle z|M(y)|x\rangle|^2 dy}{\int_Y |\langle z|M(y)|x\rangle|^2 dy}. \qquad (136)$$

If we choose the symmetric Gaussian case (135) with $\delta = 0$ and take the form of $M(y)$ without additional unitary disturbance,

$$M(y) = h \exp\left(-\frac{(y-\epsilon)^2}{4\sigma^2}\right)\Big/\sqrt{\sigma\sqrt{2\pi}}$$
$$+ v \exp\left(-\frac{(y+\epsilon)^2}{4\sigma^2}\right)\Big/\sqrt{\sigma\sqrt{2\pi}}, \qquad (137)$$

and relate both pure states to the reference state h via unitary rotations as defined in Eq. (79), $x = \mathcal{U}_{\alpha,\beta,\gamma}(h)$ and $z = \mathcal{U}_{\alpha',\beta',\gamma'}(h)$, then the postselected probability density $_z\tilde{p}_x(y)$ takes the form

$$|\langle z|M(y)|x\rangle|^2 = \frac{\exp\left(-\frac{y^2+\epsilon^2}{2\sigma^2}\right)}{2\sigma\sqrt{2\pi}}\left[(1 + \cos\beta\cos\beta')\cosh\frac{y\epsilon}{\sigma^2}\right.$$
$$+ (\cos\beta + \cos\beta')\sinh\frac{y\epsilon}{\sigma^2}$$
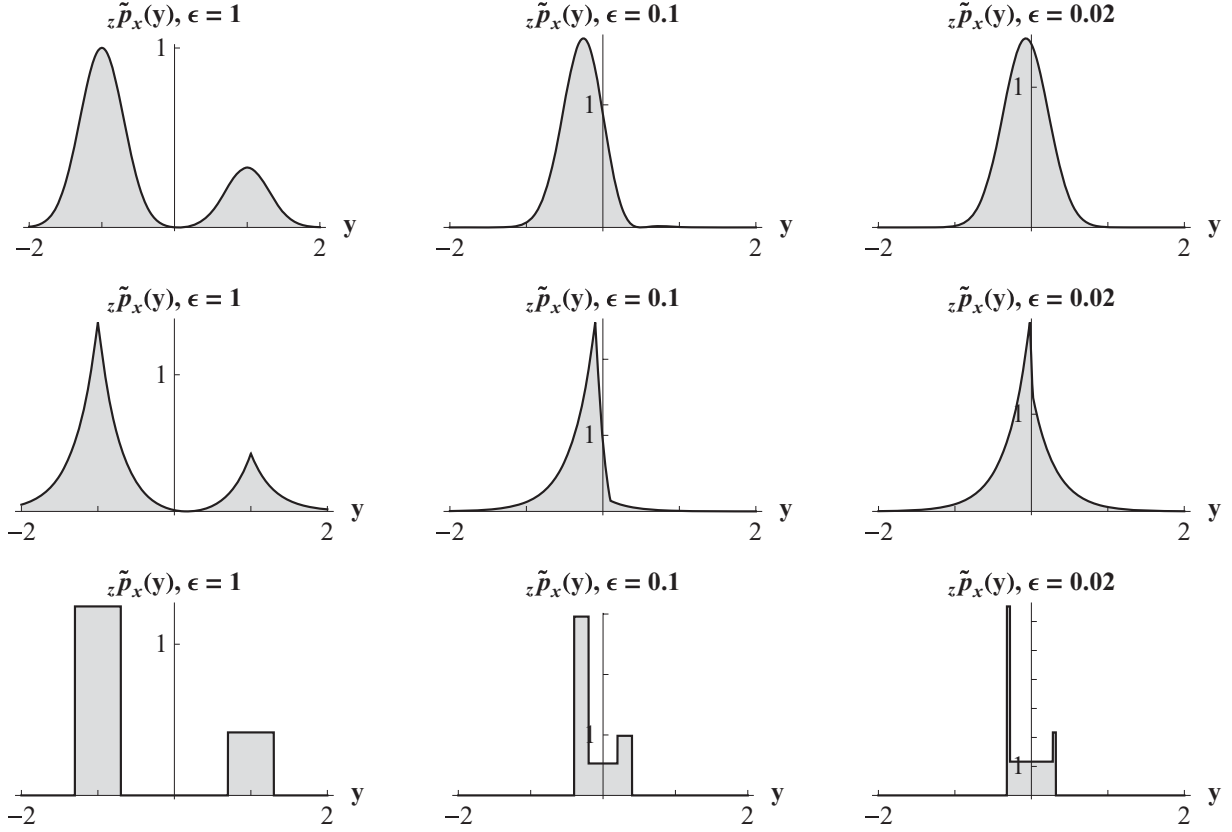$$\left.+ \sin\beta\sin\beta'\cos(\gamma - \gamma')\right]. \qquad (138)$$

FIG. 4. Pre- and postselected detector probability densities $_z\tilde{p}_x(y)$ for the calcite position measurement (131), shown for strong separation ($\epsilon = 1$), wimpy separation ($\epsilon = 0.1$), and weak separation ($\epsilon = 0.02$) of the polarizations. The preselection is $x = |x\rangle\langle x|$ with associated vector $|x\rangle = \cos(4\pi/6)|h\rangle + \sin(4\pi/6)|v\rangle$. The postselection is $z = |z\rangle\langle z|$ with associated vector $|z\rangle = (|h\rangle + |v\rangle)/\sqrt{2}$. (top row) Initial Gaussian beam profile. (middle row) Initial Laplace beam profile. (bottom row) Initial top-hat beam profile. Note that the Gaussian profile tilts to approximate a single-shifted Gaussian under weak separation, as leveraged in the weak-measurement protocol introduced in Ref. [17].

Choosing the CVs (135f) to target the observable $h - v$, the conditioned average (136) then takes the form

$$_z\widetilde{\langle h - v \rangle}_x = \frac{\cos\beta + \cos\beta'}{1 + \cos\beta\cos\beta' + \Xi(\epsilon,\sigma)}, \tag{139a}$$

$$\Xi(\epsilon,\sigma) = \sin\beta\sin\beta'\cos(\gamma - \gamma')\exp\left(-\frac{\epsilon^2}{2\sigma^2}\right). \tag{139b}$$

The interference term $\Xi(\epsilon,\sigma)$ in the denominator is the only part of the conditioned average that depends on the details of the measurement context through the exponential dependence on ϵ/σ, which was also noted in Refs. [27,91]. This conditioned average can exceed the eigenvalue range of the observable due to the combination of the amplified CV and the disturbance linking the incompatible frameworks in the conditional probabilities. Figure 5 shows the Gaussian measurement of the conditioned average (139), as well as top-hat and triangular measurements for comparison.

The conditioned average (139) has two limiting cases that eliminate the explicit context-dependence: (1) In the strong-measurement limit, $\epsilon/\sigma \to \infty$, the interference term vanishes, leaving a conditioned average of projective measurements that always stays in the eigenvalue range of the observable.

(2) In the weak-measurement limit, $\epsilon/\sigma \to 0$, the conditioned average reduces to the *weak value*,

$$
\begin{aligned}
&_z\widetilde{\langle h - v \rangle}_x^w \\
&= \mathrm{Re}\frac{\langle z|(h - v)|x\rangle}{\langle z|x\rangle} \\
&= \frac{\cos\beta + \cos\beta'}{1 + \cos\beta\cos\beta' + \sin\beta\sin\beta'\cos(\gamma - \gamma')}.
\end{aligned} \tag{140}
$$

The weak value is distinguished by being the only case that can be written entirely in terms of the observable, the postselection, and the preselected state without reference to the intermediate measurement. In this sense, it is the only context-independent form of the conditioned average. However, we shall see in Sec. III E that the weak value is not guaranteed as a limit point of the conditioned average in the weak-measurement limit.

3. Example: Three-box paradox

We can also use contextual values and the general conditioned average to analyze an often repeated paradox related to the logic of weak values: the three-box paradox [92–95]. Suppose one has three boxes, only one of which may be occupied by some quantum particle. The boxes form a classical sample space, $X = \{a, b, c\}$, with Boolean algebra $\Sigma_X = \{0, a, b, c, a + b, b + c, c + a, 1_X\}$, with $1_X = a + b +$
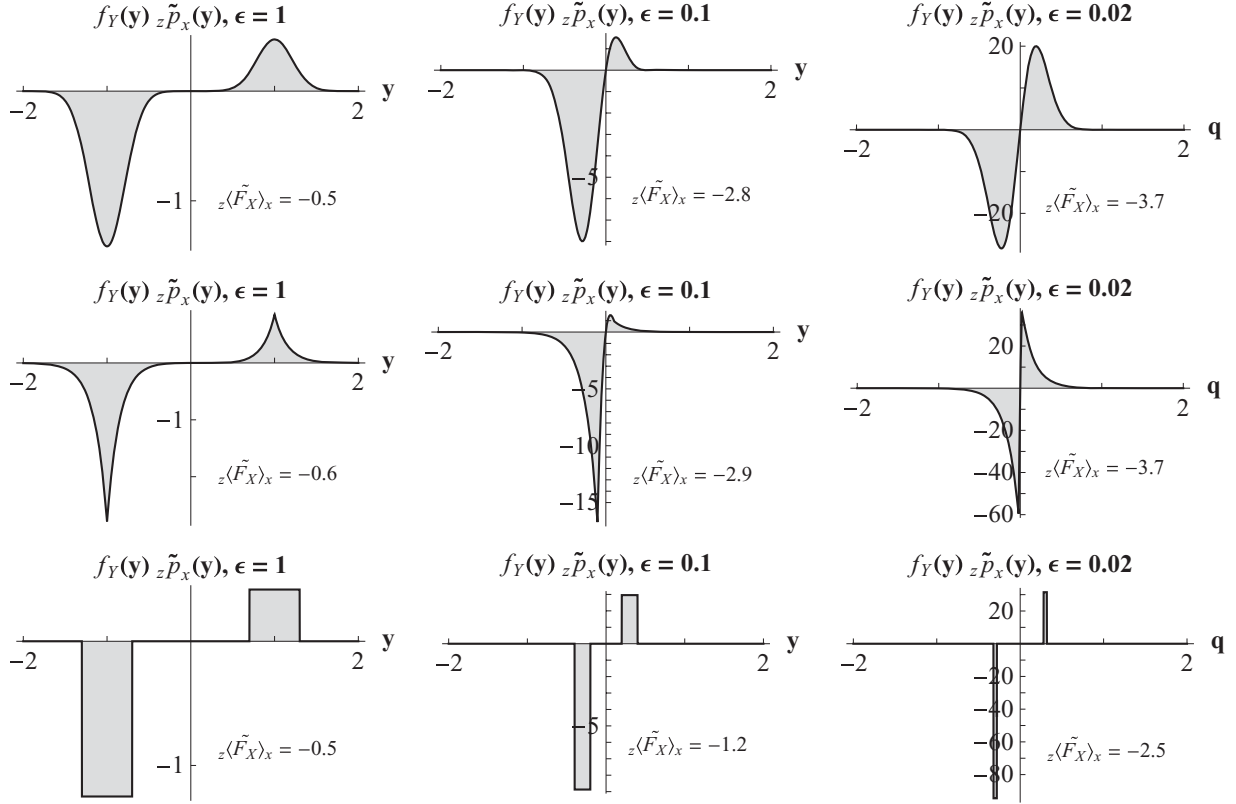
FIG. 5. Pre- and postselected conditioned average densities $f_Y(y)_z\tilde{p}_x(y)$ for a calcite position measurement targeting the observable $F_X = h - v$ with CVs as in Fig. 3, shown for strong separation ($\epsilon = 1$), wimpy separation ($\epsilon = 0.1$), and weak separation ($\epsilon = 0.02$) of the polarizations. The conditioned averages $_z\langle\widetilde{F_X}\rangle_x = \int_Y f_Y(y)_z\tilde{p}_x(y)dy$ are the areas under the curves and are shown inset. As in Fig. 4, the preselection is $x = |x\rangle\langle x|$, where $|x\rangle = \cos(4\pi/6)|h\rangle + \sin(4\pi/6)|v\rangle$. The postselection is $z = |z\rangle\langle z|$, where $|z\rangle = (|h\rangle + |v\rangle)/\sqrt{2}$. (top row) Initial Gaussian beam profile. (middle row) Initial Laplace beam profile. (bottom row) Initial top-hat beam profile. For sufficiently strong separation, all three detector profiles will produce the strong-conditioned average $_z\langle\widetilde{F_X}\rangle_x = -1/2$. For weak separation, all three profiles approximate the weak value $_z\langle\widetilde{F_X}\rangle_x^w = -2 - \sqrt{3} \approx -3.73$. However, the different detector profiles converge to the weak value at different rates with decreasing ϵ.

c. Suppose that the boxes are preselected in the pure state with density $x = |x\rangle\langle x|$ and associated Hilbert space vector $|x\rangle = (|a\rangle + |b\rangle + |c\rangle)/\sqrt{3}$ and then later postselected with the pure projector $z = |z\rangle\langle z|$ and associated vector $|z\rangle = (|a\rangle + |b\rangle - |c\rangle)/\sqrt{3}$. The postselected state has a transition probability from the preselected state of $D_x(z) = |\langle z|x\rangle|^2 = 1/9$.

According to the weak-value definition (119), the weak values of the box-occupation observables for this pre- and postselected situation are

$$_z\langle\widetilde{a}\rangle_x^w = 1, \tag{141a}$$

$$_z\langle\widetilde{b}\rangle_x^w = 1, \tag{141b}$$

$$_z\langle\widetilde{c}\rangle_x^w = -1. \tag{141c}$$

These values have occasionally been interpreted as the counterfactual conditional probabilities of box occupation given the double boundary conditions; that is, the box occupation was not checked in between the pre- and postselection, but if it *had* been checked without disturbing the system, then these probabilities would have been observed. Part of the paradox is that the weak value for *c* is negative, despite the fact that the eigenvalues for the occupation projector *c* are 1 and 0 and cannot produce such a negative conditioned average unless

negative conditional probabilities average the eigenvalues. Moreover, if the weak values do represent counterfactual probabilities, then the weak values for *a* and *b* both indicate a counterfactual certainty of occupation, and hence require a negative counterfactual probability for *c* to correctly maintain the probability normalization condition.

Operationally, the weak value is an idealized limit point of a pre- and postselected conditioned average. Since measuring it is not strictly achievable in the laboratory, we prefer to analyze this situation by considering a specific measurement context containing experimentally observable quantities. In particular, we shall consider a detector for the three-box occupation that has the three outcomes: 1, 2, and 3. The measurement operations are fully characterized by the single-measurement operators,

$$M_1 = a\sqrt{(1+\epsilon)/3} + b\sqrt{(1-\epsilon)/3} + c\sqrt{1/3}, \tag{142a}$$

$$M_2 = a\sqrt{(1-\epsilon)/3} + b\sqrt{1/3} + c\sqrt{(1+\epsilon)/3}, \tag{142b}$$

$$M_3 = a\sqrt{1/3} + b\sqrt{(1+\epsilon)/3} + c\sqrt{(1-\epsilon)/3}, \tag{142c}$$

corresponding to the probability observables $E_1 = M_1^2$, $E_2 = M_2^2$, and $E_3 = M_3^2$. For the particular pre- and postselection under consideration, these measurement

operators produce the generalized ABL conditional probabilities,

$$_z\langle\widetilde{1}\rangle_x = \frac{\text{Tr}(zM_1xM_1)}{\sum_{i=1}^3 \text{Tr}(zM_ixM_i)}$$

$$= \frac{3 - 2\sqrt{1+\epsilon} - 2\sqrt{1-\epsilon} + 2\sqrt{1-\epsilon^2}}{9 - 2\sqrt{1+\epsilon} - 2\sqrt{1-\epsilon} - 2\sqrt{1-\epsilon^2}}$$

$$= \frac{1}{3} - \frac{\epsilon^2}{3} + O(\epsilon^3), \qquad (143a)$$

$$_z\langle\widetilde{2}\rangle_x = \frac{\text{Tr}(zM_2xM_2)}{\sum_{i=1}^3 \text{Tr}(zM_ixM_i)}$$

$$= \frac{3 - 2\sqrt{1+\epsilon} + 2\sqrt{1-\epsilon} - 2\sqrt{1-\epsilon^2}}{9 - 2\sqrt{1+\epsilon} - 2\sqrt{1-\epsilon} - 2\sqrt{1-\epsilon^2}}$$

$$= \frac{1}{3} - \frac{2\epsilon}{3} + \frac{\epsilon^2}{6} + O(\epsilon^3), \qquad (143b)$$

$$_z\langle\widetilde{3}\rangle_x = \frac{\text{Tr}(zM_3xM_3)}{\sum_{i=1}^3 \text{Tr}(zM_ixM_i)}$$

$$= \frac{3 + 2\sqrt{1+\epsilon} - 2\sqrt{1-\epsilon} - 2\sqrt{1-\epsilon^2}}{9 - 2\sqrt{1+\epsilon} - 2\sqrt{1-\epsilon} - 2\sqrt{1-\epsilon^2}}$$

$$= \frac{1}{3} + \frac{2\epsilon}{3} + \frac{\epsilon^2}{6} + O(\epsilon^3). \qquad (143c)$$

These detection probabilities are all positive and well formed, since they are operationally accessible quantities.

If we target a particular observable $O_X = o_X(a)a + o_X(b)b + o_X(c)c$ for the three boxes, we can solve for the appropriate CV by inverting the matrix equation

$$\begin{pmatrix} o_X(a) \\ o_X(b) \\ o_X(c) \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 1+\epsilon & 1-\epsilon & 1 \\ 1-\epsilon & 1 & 1+\epsilon \\ 1 & 1+\epsilon & 1-\epsilon \end{pmatrix}\begin{pmatrix} o_Y(1) \\ o_Y(2) \\ o_Y(3) \end{pmatrix}, \qquad (144)$$

producing

$$\begin{pmatrix} o_Y(1) \\ o_Y(2) \\ o_Y(3) \end{pmatrix} = \frac{o_X(a) + o_X(b) + o_X(c)}{3} + \frac{1}{\epsilon}\begin{pmatrix} o_X(a) - o_X(b) \\ o_X(c) - o_X(a) \\ o_X(b) - o_X(c) \end{pmatrix}. \qquad (145)$$

In particular, we can use these CVs to expand the box-occupation observables in terms of the probability observables,

$$a = \left(\frac{1}{3} + \frac{1}{\epsilon}\right)E_1 + \left(\frac{1}{3} - \frac{1}{\epsilon}\right)E_2 + \frac{1}{3}E_3$$

$$= \frac{1}{3}1_X + \frac{1}{\epsilon}(E_1 - E_2), \qquad (146a)$$

$$b = \left(\frac{1}{3} - \frac{1}{\epsilon}\right)E_1 + \frac{1}{3}E_2 + \left(\frac{1}{3} + \frac{1}{\epsilon}\right)E_3$$

$$= \frac{1}{3}1_X + \frac{1}{\epsilon}(E_3 - E_1), \qquad (146b)$$

$$c = \frac{1}{3}E_1 + \left(\frac{1}{3} + \frac{1}{\epsilon}\right)E_2 + \left(\frac{1}{3} - \frac{1}{\epsilon}\right)E_3$$

$$= \frac{1}{3}1_X + \frac{1}{\epsilon}(E_2 - E_3). \qquad (146c)$$

Hence, all three box-occupation observables can be measured simultaneously from the same set of probabilities for the

three detector outcomes. Notably, the CV assigned to each outcome can be *negative* for sufficiently small ϵ, even though all eigenvalues are positive or zero. Hence the values being averaged can be negative and thus can lead to negative averages in principle.

Computing the appropriate conditioned averages we find to $O(\epsilon^3)$,

$$_z\langle\widetilde{a}\rangle_x = 1 - \frac{\epsilon}{2} - \frac{\epsilon^2}{4} + O(\epsilon^3), \qquad (147a)$$

$$_z\langle\widetilde{b}\rangle_x = 1 + \frac{\epsilon}{2} - \frac{\epsilon^2}{4} + O(\epsilon^3), \qquad (147b)$$

$$_z\langle\widetilde{c}\rangle_x = -1 + \frac{\epsilon^2}{2} + O(\epsilon^3), \qquad (147c)$$

which shows that the weak values (141) are the $\epsilon \to 0$ limit of the conditioned averages with this specific measurement context.

The paradox of the negative weak value (141) can therefore be largely resolved in the following sense: the combination of the amplified negative CVs and the disturbance in the detector probabilities linking pre- and postselection frameworks leads to the negative result for $_z\langle\widetilde{c}\rangle_x$ given sufficiently small ϵ. No negative probabilities are required to obtain the negative limit point since negative CVs are being averaged in the weak limit and not eigenvalues. All operationally accessible probabilities are positive and well behaved: the negative CVs are assigned by the experimenter and highlighted by the disturbance in the well-behaved probabilities.

We leave the reader to ponder how to interpret the operationally accessible negative conditioned average (147c). However, we note that, with at least this measurement context, the conditioned averages do obey the equality

$$_z\langle\widetilde{a}\rangle_x + {}_z\langle\widetilde{b}\rangle_x + {}_z\langle\widetilde{c}\rangle_x = 1, \qquad (148)$$

for all values of ϵ. The three sets of CVs sum to unity for each detector outcome, leaving only the normalized sum of detector probabilities $_z\langle\widetilde{1}\rangle_x + {}_z\langle\widetilde{2}\rangle_x + {}_z\langle\widetilde{3}\rangle_x = 1$. For more discussion of this paradox, see, for example, [92–95].

E. Deriving the weak value

Weak-value controversy. As we have seen for the case of the calcite detector (140) and the three-box paradox (141), the weak value (118) seems to arise naturally as the weak limit of postselected conditioned averages. Indeed, much of the existing literature on weak values (e.g., [17,19–21,24–26]) operates under the assumption that it is the only weak limit of a conditioned average, or that it is a well-defined property of a pre- and postselected ensemble prior to the ensemble being measured. However, a conditioned average does not necessarily converge to the weak value in the weak-measurement limit, as has been noted independently by several groups [22,23,27,32,35,49,96], making its interpretation as a well-defined property worthy of more careful consideration. To obtain correct laboratory predictions for a conditioned average, the formula (115) must be used, which generally requires the specification of the detection strategy and the protocol for assigning CVs to target a specific observable.

Despite the interpretational controversy, the weak value (118) is distinguished by being a *context-independent* weak limit of the conditioned average that is easy to compute theoretically and appears quite commonly in typical laboratory situations. The formal expression of the weak value can also appear in other measurement scenarios, such as in "modular values" [97], or even perturbative corrections to energy spectra [98], which makes it an independently interesting quantity to study.

We will now demonstrate how the weak value (118) can be uniquely defined from the general conditioned average (115) by imposing a set of sufficient conditions that the measurement should satisfy.

Preliminaries. First we note from (95c) that each measurement operator has a polar decomposition, $M_{y,y'} = U_{y,y'}|M|_{y,y'}$, in terms of a unitary operator $U_{y,y'}$ and a positive operator $|M|_{y,y'}$. It then follows that

$$
\begin{aligned}
M_{y,y'}^{\dagger} E_z' M_{y,y'} &= |M|_{y,y'} U_{y,y'}^{\dagger} E_z' U_{y,y'} |M|_{y,y'} \\
&= \{|M|_{y,y'}^2, \mathcal{U}_{y,y'}(E_z')\}/2 \\
&\quad - [|M|_{y,y'},[|M|_{y,y'},\mathcal{U}_{y,y'}(E_z')]]/2, \quad (149)
\end{aligned}
$$

where $\{A,B\} = AB + BA$ is the anticommutator, $[A,B] = AB - BA$ is the commutator, and $\mathcal{U}_{y,y'}(E_z') = U_{y,y'}^{\dagger} E_z' U_{y,y'}$ is a unitary rotation of the postselection.

Sufficient conditions. Next we make the following sufficient assumptions regarding the dependence of the relevant quantities on the measurement strength parameter ϵ:

(1) The measurement operators $M_{y,y'}$ are analytic functions of ϵ, and thus have well-defined Taylor expansions around $\epsilon = 0$ such that they are proportional to the identity in the weak limit, $\forall\, y, y', \lim_{\epsilon \to 0} M_{y,y'} \propto 1_X$.

(2) The unitary parts of the measurement operators $U_{y,y'} = \exp(iG_{y,y'}(\epsilon))$ are generated by Hermitian operators of order ϵ^k, $G_{y,y'}(\epsilon) = \epsilon^k G_{y,y'}^{(k)} + O(\epsilon^{k+1})$, for some integer $k \geqslant 1$. Furthermore, each $U_{y,y'}$ must commute with either the system state or the postselection, $\forall\, y, y', [U_{y,y'}, \rho_X] = 0$, or $\forall\, y, y', [U_{y,y'}, E_z'] = 0$.

(3) The equality $F_X = \sum_y f_Y(\epsilon; y) E_y(\epsilon)$ must be satisfied, where the CVs $f_Y(\epsilon; y)$ are selected according to the pseudoinverse prescription.

(4) The minimum nonzero order in ϵ for all $|M|_{y,y'}(\epsilon)$ is ϵ^n such that (3) can also be satisfied for some CVs by the truncation to order ϵ^n. That is, for all y, y', $|M|_{y,y'} = c_{y,y'}1_X + |M|_{y,y'}^{(n)}\epsilon^n + O(\epsilon^{n+1})$, where $\sum_{y'} c_{y,y'}^2 = P_Y(y)$ is the detector probability in the absence of interaction, and some of the $|M|_{y,y'}^{(n)}$ may vanish.

(5) The probability observables $E_y(\epsilon) = \sum_{y'} M_{y,y'}^{\dagger}(\epsilon) M_{y,y'}(\epsilon)$ commute with the observable F_X.

Theorem. Given the above sufficient conditions, we have the following theorem: In the weak limit $\epsilon \to 0$, the context dependence of the conditioned average (115) vanishes and the weak value (118) is uniquely defined.

Proof. To prove the theorem, we expand (115) to the minimum necessary order of ϵ^n and then take the weak limit as $\epsilon \to 0$. First, we expand (149) to order ϵ^n using assumptions

(1) and (4),

$$
\begin{aligned}
M_{y,y'}^{\dagger} E_z' M_{y,y'} &= c_{y,y'}^2 \mathcal{U}_{y,y'}(E_z') + c_{y,y'}\{|M|_{y,y'}^{(n)}, \mathcal{U}_{y,y'}(E_z')\}\epsilon^n + O(\epsilon^{n+1}).
\end{aligned}
\tag{150}
$$

Generally, the remaining unitary rotation of the postselection will disturb the weak limit. However, if $[U_{y,y'}, E_z'] = 0$ as in assumption (2), then $\mathcal{U}_{y,y'}(E_z') = E_z'$ and the unitary disturbance disappears. If instead $[U_{y,y'}, \rho_X] = 0$, then we can apply the state to (150) and find

$$
\begin{aligned}
\langle M_{y,y'}^{\dagger} E_z' M_{y,y'} \rangle_X &= c_{y,y'}^2 \langle \mathcal{U}_{y,y'}(E_z') \rangle_X + c_{y,y'} \langle \{|M|_{y,y'}^{(n)}, \mathcal{U}_{y,y'}(E_z')\} \rangle_X \epsilon^n \\
&\quad + O(\epsilon^{n+1}).
\end{aligned}
\tag{151}
$$

Since $\langle \mathcal{U}_{y,y'}(E_z') \rangle_X = \mathrm{Tr}_X(\mathcal{U}^{\dagger}(\rho_X) E_z') = \langle E_z' \rangle_X$, the first term simplifies. The unitary rotation in the second term expands to $\mathcal{U}_{y,y'}(E_z') = E_z' + O(\epsilon^k)$, and the $O(\epsilon^k)$ correction can be absorbed into the overall $O(\epsilon^{n+1})$ correction.

Therefore, after summing over y' we find, up to corrections of order ϵ^{n+1},

$$
\sum_{y'} \langle M_{y,y'}^{\dagger} E_z' M_{y,y'} \rangle_X = \langle \{E_y(\epsilon), E_z'\} \rangle_X / 2, \tag{152}
$$

where the probability observable has the expansion to order ϵ^n of

$$
\begin{aligned}
E_y(\epsilon) &= \sum_{y'} |M|_{y,y'}^2(\epsilon) \\
&= \sum_{y'} \left[c_{y,y'}^2 1_X + 2c_{y,y'}|M|_{y,y'}^{(n)}\epsilon^n + O(\epsilon^{n+1}) \right]. \tag{153}
\end{aligned}
$$

Inserting (152) into (115), we find

$$
{}_z\langle \widetilde{F_X} \rangle = \frac{\langle \{F_X, E_z'\}/2 \rangle_X + \sum_y f_Y(\epsilon; y)O(\epsilon^{n+1})}{\langle \{1_X, E_z'\}/2 \rangle_X + O(\epsilon^{n+1})}, \tag{154}
$$

where we have simplified $\sum_y f_Y(\epsilon; y) E_y(\epsilon) = F_X$ in the numerator, and $\sum_y E_y(\epsilon) = 1_X$ in the denominator. Hence, unless the CVs in the numerator have poles larger than $1/\epsilon^n$, the correction terms of order ϵ^{n+1} will vanish, producing (118) in the weak limit $\epsilon \to 0$, as claimed. The last step in obtaining (118), therefore, is to show that the pseudoinverse solution for f_Y that was indicated by assumption (3) cannot have poles larger than $1/\epsilon^n$. The following lemmas will show this, which will prove the main theorem.

Lemma preliminaries. We note that F_X commutes with $\{E_y(\epsilon)\}$ by assumption (5). As such, we will replace the CV definition $F_X = \sum_y f_Y(\epsilon; y) E_y(\epsilon)$ with an equivalent matrix equation,

$$
\vec{f}_X = \mathcal{S} \vec{f}_Y, \tag{155a}
$$

$$
\mathcal{S} = \begin{pmatrix} \langle E_y(\epsilon) \rangle_x & \cdots \\ \vdots & \ddots \end{pmatrix}. \tag{155b}
$$

The pseudoinverse is constructed from the singular value decomposition $\mathcal{S} = \mathcal{U}\Sigma\mathcal{V}^T$ as $\mathcal{S}^+ = \mathcal{V}\Sigma^+\mathcal{U}^T$, where $\mathcal{U}$ and $\mathcal{V}$ are orthogonal matrices such that $\mathcal{U}^T\mathcal{U} = \mathcal{V}\mathcal{V}^T = 1$, Σ is the singular value matrix composed of the square roots of

the eigenvalues of SS^T, and Σ^+ is composed of the inverse nonzero elements in Σ^T.

Next, we note that the truncation of the matrix S to order ϵ^n has the form,

$$S' = \mathcal{P} + \epsilon^n S_n, \qquad (156)$$

where $\mathcal{P} = (P_Y(y)\vec{1}, \cdots)$ is a matrix whose rows are identical and whose columns contain the interaction-free detector probabilities $P_Y(y)$, and $S_n = (\vec{E}_1^{(n)}, \cdots)$ is a matrix whose rows all sum to zero. Furthermore, since the solution to the equation $\vec{f}_X = S'\vec{f}_Y'$ is assumed to exist by assumption (4), then the dimension of the detector, N, must be greater than or equal to the dimension of the system, M. We then have the following two lemmas.

Lemma 1. The singular values of the truncated matrix S' have maximum leading order ϵ^n.

Proof. The singular values of S' are $\sigma_k = \sqrt{\lambda_k}$, where λ_k are M eigenvalues of $\mathcal{H} = S^T S$, with its other $N - M$ eigenvalues being zero. Since $\mathcal{P}^T S_n = 0$, this matrix has the simple form $\mathcal{H} = \mathcal{P}^T\mathcal{P} + \epsilon^{2n}S_n^T S_n$, where $(\mathcal{P}^T\mathcal{P})_{ij} = M P_Y(i)P_Y(j)$ is $M||\vec{p}||^2$ times the projection operator onto the probability vector $\vec{p} = (P_Y(y), \cdots)$, and $(S_n^T S_n)_{ij} = \vec{E}_i^{(n)} \cdot \vec{E}_j^{(n)}$. We will use $\mathcal{H}$ to determine the singular values of S'.

Differentiating the eigenvalue relation $\mathcal{H}(\epsilon^{2n})\vec{u}_k(\epsilon^{2n}) = \lambda_k(\epsilon^{2n})\vec{u}_k(\epsilon^{2n})$ and the eigenvector normalization condition $\vec{u}_k(\epsilon^{2n}) \cdot \vec{u}_k(\epsilon^{2n}) = 1$ with respect to ϵ^{2n} produces the following deformation equation that describes how the eigenvalues of $\mathcal{H}$ continuously change with increasing ϵ^{2n},

$$\dot{\lambda}_k(\epsilon^{2n}) = ||S_n\vec{u}_k(\epsilon^{2n})||^2. \qquad (157)$$

Integrating this equation produces the following perturbative expansion of the eigenvalues for small ϵ,

$$\lambda_k(\epsilon^{2n}) = \lambda_k(0) + \epsilon^{2n}||S_n\vec{u}_k(0)||^2 + O(\epsilon^{4n}). \quad (158)$$

Hence, to prove the lemma it is sufficient to show that $\lambda_k(0)$ and $S_n\vec{u}_k(0)$ cannot both vanish unless $\lambda_k(\epsilon^{2n}) = 0$ for all ϵ.

Since $\mathcal{H}(0) = \mathcal{P}^T\mathcal{P}$ is a projection operator, $\lambda_1(0) = M||\vec{p}||^2$ is its only nonzero eigenvalue with associated eigenvector $\vec{u}_1(0) = \vec{p}/||\vec{p}||$. Hence, $\sigma_1(\epsilon^{2n}) \approx \sqrt{M}||\vec{p}|| > 0$ to leading order. For $k \neq 1$, $\lambda_k(0) = 0$ and $\vec{u}_k(0)$ can be chosen arbitrarily to span the degenerate $(N-1)$-dimensional subspace orthogonal to $\vec{u}_1(0)$. Suppose $S_n\vec{u}_k(0) = 0$ for some k. It follows that $\mathcal{H}(\epsilon^{2n})\vec{u}_k(0) = \mathcal{P}^T\mathcal{P}\vec{u}_k(0) + \epsilon^{2n}S_n^T S_n\vec{u}_k(0) = 0$ since $\vec{u}_k(0)$ is orthogonal to $\vec{u}_1(0) \propto \vec{p}$. Therefore, $\vec{u}_k(0)$ is an eigenvector of $\mathcal{H}(\epsilon^{2n})$ with eigenvalue 0 for any ϵ. Since $\mathcal{H}$ is symmetric, its eigenvectors form an orthogonal set for any ϵ, so we must have the identification $\vec{u}_k(\epsilon^{2n}) = \vec{u}_k(0)$. As a result, the associated eigenvalue vanishes for any ϵ, $\lambda_k(\epsilon^{2n}) = \lambda_k(0) = 0$, which proves the lemma.

Lemma 2. The pseudoinverse solution $\vec{f}_Y$ to (155a) cannot have poles larger than $1/\epsilon^n$.

Proof. In order to satisfy (155a), we have the equivalent condition for each component of $\mathcal{U}^T\vec{f}_X = \Sigma\mathcal{V}^T\vec{f}_Y$,

$$(\mathcal{U}^T\vec{f}_X)_k = \Sigma_{kk}(\mathcal{V}^T\vec{f}_Y)_k. \qquad (159)$$

Therefore, all singular values Σ_{kk} corresponding to nonzero components of $\mathcal{U}^T\vec{f}_X$ must also be nonzero; we shall call these the *relevant* singular values. Singular values which

are not relevant will not contribute to the solution $\vec{f}_Y = \mathcal{V}\Sigma^+\mathcal{U}^T\vec{f}_X$. We can see this since $(\vec{f}_Y)_j = (\mathcal{V}\Sigma^+\mathcal{U}^T\vec{f}_X)_j = \sum_k \mathcal{V}_{jk}\Sigma_{kk}^+(\mathcal{U}^T\vec{f}_X)_k$, so any zero element of $\mathcal{U}^T\vec{f}_X$ will eliminate the inverse irrelevant singular value Σ_{kk}^+ from the solution for $(\vec{f}_Y)_j$.

Since the orthogonal matrices $\mathcal{U}$ and $\mathcal{V}$ do not contain any poles, and since $\vec{f}_X$ is ϵ independent, then the only poles in the solution $\vec{f}_Y = S^+\vec{f}_X = \mathcal{V}\Sigma^+\mathcal{U}^T\vec{f}_Y$ must come from the inverses of the relevant singular values in Σ^+. If a singular value Σ_{kk} has leading order ϵ^m, then its inverse $\Sigma_{kk}^+ = 1/\Sigma_{kk}$ has leading order $1/\epsilon^m$; therefore, to have a pole of order higher than $1/\epsilon^n$ then there must be at least one relevant singular value with a leading order greater than ϵ^n. However, if that were the case then the truncation S' of S to order ϵ^n could not satisfy (159) since, to that order, it would have a relevant singular value of zero according to the previous lemma, contradicting assumption (4) about needing to satisfy the CV definition with the minimum nonzero order in ϵ. Therefore, the pseudoinverse solution $\vec{f}_Y = S^+\vec{f}_X$ can have no pole with order higher than $1/\epsilon^n$ and the theorem is proven.

Exceptions. As the theorem indicates, the weak value will arise as the weak limit of a conditioned average in many common laboratory situations, which explains its seeming stability in the literature. However, if the sufficiency conditions of the theorem are not met, then a different weak limit may be found. For example, if there is ϵ-dependent unitary disturbance in the measurement, then the postselection can be effectively rotated to a different framework for each measurement outcome, which creates additional terms in the weak limit. (See, for example, Ref. [35].) Similarly, if the CVs are ϵ-dependent and diverge more rapidly than $1/\epsilon^n$, then additional terms will become relevant in the weak limit. This latter case can happen either from a pathological choice of CV by the experimenter in the case of redundancy, or from a set of probability observables that cannot satisfy the constraint equation $F_X = \sum_y f_Y(\epsilon; y)E_y(\epsilon)$ with their lowest nonzero order in ϵ. Such probability observables that do not satisfy the constraint equation to lowest order are poorly correlated with the observable in the weak limit. We refer the reader to [53] for more discussion on the uniqueness issue of weak values. The theorem presented here is a slight generalization of the one presented therein.

IV. CONCLUSION

In this work, we have detailed the contextual-value approach to the generalized measurement of observables that we originally introduced in the letter [49] and further developed in Refs. [51–53]. This approach completes the well-established operational theory of state measurements by directly relating the state transformations to traditional observables. Each such operation typically corresponds to a distinguishable outcome of a correlated detection apparatus. An experimenter can construct an observable from such an apparatus by assigning values to its outcomes. The assigned values can be generally amplified from the eigenvalues of the constructed observable due to ambiguity in the measurement, and thus form a generalized spectrum that depends on the specific

measurement context. Hence, we call these values *contextual values* for the constructed observable that allow its indirect measurement using such a correlated detector.

Constructing an observable using contextual values requires only classical probability theory, according to (32). Hence, the technique may be used wherever Bayesian probability theory applies. We have outlined an algebraic approach to operational measurements from within Bayesian probability theory to encourage applications along these lines.

We have also shown how to construct a quantum probability space as the orbit of a classical probability space under the special unitary group. This point of view illustrates that quantum observables can be constructed from contextual values in precisely the same way (112) as their classical counterparts. The approach also highlights the similarity between Lüder's rule (89) for updating a quantum state and *invasive* classical conditioning (13), which leads to a similarity between quantum operations (95) and classically invasive measurement operations (25). Numerous physical examples have been given.

By putting all observable measurements on the same footing, the contextual-value formalism subsumes not only projective measurements but also weak measurements as special cases. To emphasize this point, we have analyzed the quantum weak measurement protocol introduced by Aharonov *et al.* [17] in detail as an example using a calcite crystal and a polarized laser beam. We have also derived the quantum weak value (118) as a limit point of a general pre- and postselected conditioned average (115) as the measurement strength goes to zero and have given sufficient conditions for the convergence to hold. Like the classically invasive conditioned average (37), the quantum conditioned average, with the quantum weak value as a special case, can exceed the eigenvalue bounds of the observable.

The use of contextual values considerably clarifies and formalizes the process of measuring observables, particularly within a laboratory setting. The elements of the formalism directly describe operationally accessible quantities that can be tomographically calibrated. As such, the technique should be of considerable interest to experimentalists working on measurement and control of both quantum and classical systems. Furthermore, the formalism prompts interesting theoretical questions about the foundations of quantum mechanics by highlighting its myriad similarities to classical probability theory.

ACKNOWLEDGMENTS

[1] P. A. M. Dirac, *Principles of Quantum Mechanics* (Oxford University Press, Oxford, 1930).

[2] J. von Neumann, *Mathematische Grundlagen der Quantenmechanik* (Springer, Berlin, 1932).

[3] A. N. Korotkov and A. N. Jordan, Phys. Rev. Lett. **97**, 166805 (2006).

[4] N. Katz *et al.*, Phys. Rev. Lett. **101**, 200401 (2008).

[5] Y.-S. Kim, Y.-W. Cho, Y.-S. Ra, and Y.-H. Kim, Opt. Express **17**, 11978 (2009).

[6] E. B. Davies and J. T. Lewis, Commun. Math. Phys. **17**, 239 (1970).

[7] K. Kraus, Annals of Physics **64**, 311 (1971).

[8] K. Kraus, *States, Effects and Operations: Fundamental Notions of Quantum Theory* (Springer-Verlag, Berlin, 1983).

[9] V. Braginski and F. Khalili, *Quantum Measurement* (Cambridge University Press, Cambridge, 1992).

[10] P. Busch, M. Grabowski, and P. J. Lahti, *Operational Quantum Physics* (Springer-Verlag, Berlin, 1995).

[11] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).

[12] R. Alicki and M. Fannes, *Quantum Dynamical Systems* (Oxford University Press, Oxford, 2001).

[13] M. Keyl, Phys. Rep. **369**, 431 (2002).

[14] H. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (Oxford University Press, Oxford, 2007).

[15] S. M. Barnett, *Quantum Information* (Oxford University Press, Oxford, 2009).

[16] H. M. Wiseman and G. Milburn, *Quantum Measurement and Control* (Cambridge University Press, Cambridge, 2010).

[17] Y. Aharonov, D. Z. Albert, and L. Vaidman, Phys. Rev. Lett. **60**, 1351 (1988).

[18] I. M. Duck, P. M. Stevenson, and E. C. G. Sudarshan, Phys. Rev. D **40**, 2112 (1989).

[19] Y. Aharonov and L. Vaidman, Phys. Rev. A **41**, 11 (1990).

[20] Y. Aharonov and A. Botero, Phys. Rev. A **72**, 052111 (2005).

[21] Y. Aharonov and L. Vaidman, Lect. Notes Phys. **734**, 399 (2008).

[22] R. Jozsa, Phys. Rev. A **76**, 044103 (2007).

[23] A. Di Lorenzo and J. C. Egues, Phys. Rev. A **77**, 042108 (2008).

[24] Y. Aharonov, S. Popescu, J. Tollaksen, and L. Vaidman, Phys. Rev. A **79**, 052110 (2009).

[25] Y. Aharonov, S. Popescu, and J. Tollaksen, Phys. Today **63**, 27 (2010).

[26] J. Tollaksen, J. Phys. A **40**, 9033 (2007).

[27] T. Geszti, Phys. Rev. A **81**, 044102 (2010).

[28] A. Hosoya and Y. Shikano, J. Phys. A **43**, 385307 (2010).

[29] Y. Shikano and A. Hosoya, J. Phys. A **43**, 025304 (2010).

[30] Y. Shikano and A. Hosoya, Physica E **43**, 776 (2011).

[31] S. Kagami, Y. Shikano, and K. Asahi, Physica E **43**, 761 (2011).

[32] S. Wu and Y. Li, Phys. Rev. A **83**, 052106 (2011).

[33] E. Haapasalo, P. Lahti, and J. Schultz, e-print arXiv:1108.3663 (to be published).

[34] N. W. M. Ritchie, J. G. Story, and R. G. Hulet, Phys. Rev. Lett. **66**, 1107(1991).

[35] H. M. Wiseman, Phys. Rev. A **65**, 032111 (2002).

[36] H. M. Wiseman, Phys. Lett. A **311**, 285 (2003).

[37] N. Brunner, V. Scarani, M. Wegmuller, M. Legre, and N. Gisin, Phys. Rev. Lett. **93**, 203902 (2004).

[38] G. J. Pryde, J. L. O'Brien, A. G. White, T. C. Ralph, and H. M. Wiseman, Phys. Rev. Lett. **94**, 220405 (2005).

[39] R. Mir, J. S. Lundeen, M. W. Mitchell, A. M. Steinberg, J. L. Garretson, and H. M. Wiseman, New J. Phys. **9**, 287 (2007).

[40] O. Hosten and P. Kwiat, Science **319**, 787 (2008).

[41] P. B. Dixon, D. J. Starling, A. N. Jordan, and J. C. Howell, Phys. Rev. Lett. **102**, 173601 (2009).

[42] J. S. Lundeen and A. M. Steinberg, Phys. Rev. Lett. **102**, 020404 (2009).

[43] D. J. Starling, P. B. Dixon, A. N. Jordan, and J. C. Howell, Phys. Rev. A **80**, 041803 (2009).

[44] J. C. Howell, D. J. Starling, P. B. Dixon, P. K. Vudyasetu, and A. N. Jordan, Phys. Rev. A **81**, 033813 (2010).

[45] D. J. Starling, P. B. Dixon, A. N. Jordan, and J. C. Howell, Phys. Rev. A **82**, 063822 (2010).

[46] D. J. Starling, P. B. Dixon, N. S. Williams, A. N. Jordan, and J. C. Howell, Phys. Rev. A **82**, 011802(R) (2010).

[47] 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).

[48] S. Kocsis, B. Braverman, S. Ravets, M. J. Stevens, R. P. Mirin, L. K. Shalm, and a. M. Steinberg, Science **332**, 1170 (2011).

[49] J. Dressel, S. Agarwal, and A. N. Jordan, Phys. Rev. Lett. **104**, 240401 (2010).

[50] We acknowledge, however, that the seeds of the idea were implicitly present in earlier works [23,35,38].

[51] J. Dressel, C. J. Broadbent, J. C. Howell, and A. N. Jordan, Phys. Rev. Lett. **106**, 040402 (2011).

[52] J. Dressel, Y. Choi, and A. N. Jordan, Phys Rev. B **85**, 045320 (2012).

[53] J. Dressel and A. N. Jordan, J. Phys. A: Math. Theor. **45**, 015304 (2012).

[54] J. Bell, Rev. Mod. Phys. **38**, 447 (1966).

[55] S. Kochen and E. P. Specker, J. Math. Mech. **17**, 59 (1967).

[56] N. D. Mermin, Rev. Mod. Phys. **65**, 803 (1993).

[57] R. W. Spekkens, Phys. Rev. A **71**, 052108 (2005).

[58] M. S. Leifer and R. W. Spekkens, Phys. Rev. Lett. **95**, 200405 (2005).

[59] R. W. Spekkens, Phys. Rev. Lett. **101**, 020401 (2008).

[60] R. T. Cox, Am. J. Phys. **14**, 1 (1946).

[61] R. T. Cox, *The Algebra of Probable Inference* (John Hopkins University Press, Baltimore, 1961).

[62] E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, 2003).

[63] C. M. Caves, C. A. Fuchs, and R. Schack, Phys. Rev. A **65**, 022305 (2002).

[64] C. A. Fuchs, e-print arXiv:1003.5209 (to be published).

[65] R. W. Spekkens, Phys. Rev. A **75**, 032110 (2007).

[66] N. Harrigan and R. W. Spekkens, Found. Phys. **40**, 125 (2010).

[67] M. S. Leifer and R. W. Spekkens, e-print arXiv:1107.5849 (to be published).

[68] The natural mathematical tools to use for a more rigorous construction of what follows would be measure theory [72,73,75,76] and its noncommutative extension to von Neumann (W*) algebras [2,12,70,82,99–102].

[69] In measure theory [72–76] Σ_X is known as a σ algebra over the set X, which also motivates the notation. With the set representation, Σ_X is a nonempty set of subsets of X that is closed under countable unions (OR), countable intersections (AND), and set complementation (NOT).

[70] J. M. Jauch, *Foundations of Quantum Mechanics* (Addison-Wesley Pub. Co, Reading, 1968).

[71] The quantity $dP/d\mu$ is also known as the *Radon-Nikodym* derivative [75,99,100,102] of the measure P with respect to the measure μ.

[72] A. Robinson, *Non-Standard Analysis* (Princeton University Press, 1966).

[73] E. Nelson, *Radically Elementary Probability Theory* (Princeton University Press, 1987).

[74] R. C. Jeffrey, *The Logic of Decision* (The University of Chicago Press, 1965).

[75] W. Rudin, *Real and Complex Analysis* (McGraw Hill, 1987).

[76] P. Szekeres, *A Course in Modern Mathematical Physics: Groups, Hilbert Spaces, and Differential Geometry* (Cambridge University Press, 2004).

[77] R. B. Griffiths, e-print arXiv:1105.3932 (to be published).

[78] D. Hestenes and G. Sobczyk, *Clifford Algebra to Geometric Calculus: A Unified Language for Mathematics and Physics* (D. Riedel Publishing Co., Dordrecht, 1987).

[79] D. Hestenes, *New Foundations for Classical Mechanics* (Kluwer Academic Publishers, Dordrecht, 1999).

[80] C. Doran and A. Lasenby, *Geometric Algebra for Physicists* (Cambridge University Press, 2007).

[81] See [80] for discussion on a purely real representation of the relativistic quantum electron to drive home this point.

[82] M. Takesaki, *Theory of Operator Algebras I, II, and III* (Springer, Berlin, 1979).

[83] M. Born, Z. Phys. **37**, 863 (1926).

[84] A. Gleason, Indiana Univ. Math. J. **6**, 885 (1957).

[85] G. Lüders, Ann. Phys. **15**, 663 (2006).

[86] Y. Aharonov, P. G. Bergmann, and J. L. Lebowitz, Phys. Rev. B **134**, 1410 (1964).

[87] S. Massar and S. Popescu, Phys. Rev. A **84**, 052106 (2011).

[88] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Rev. Mod. Phys. **81**, 865 (2009).

[89] G. Lindblad, Commun. Math. Phys. **48**, 119 (1976).

[90] S. L. Braunstein and C. M. Caves, Found. Phys. Lett. **1**, 3 (1988).

[91] N. S. Williams and A. N. Jordan, Phys. Rev. Lett. **100**, 026804 (2008).

[92] Y. Aharonov and L. Vaidman, J. Phys. A **24**, 2315 (1991).

[93] K. J. Resch, J. S. Lundeen, and A. M. Steinberg, Phys. Lett. A **324**, 125 (2004).

[94] T. Ravon and L. Vaidman, J. Phys. A **40**, 2873 (2007).

[95] N. Aharon and L. Vaidman, Phys. Rev. A **77**, 052310 (2008).

[96] S. Parrott, e-print arXiv:0909.0295v3 (to be published).

[97] Y. Kedem and L. Vaidman, Phys. Rev. Lett. **105**, 230401 (2010).

[98] If $|E\rangle$ is an eigenstate of a Hamiltonian H with energy E, and $|E'\rangle$ is an eigenstate of the perturbed Hamiltonian $H' = H + \Delta$ with energy E', then $\langle E'|H'|E\rangle = E'\langle E'|E\rangle = E\langle E'|E\rangle + \langle E'|\Delta|E\rangle$, so we can conclude that the perturbation in the eigenenergy corresponds to the (real) weak value of the perturbation $E' = E + \langle E'|\Delta|E\rangle/\langle E'|E\rangle$.

[99] S. Sakai, Bull. Amer. Math. Soc. **71**, 149 (1965).

[100] H. Kosaki, Pub. Res. Inst. Math. Sci. **18**, 379 (1982).

[101] A. Connes, *Noncommutative Geometry* (Academic Press, Boston, 1994).

[102] M. Raginsky, J. Math. Phys. **44**, 5003 (2003).