TOPICAL REVIEW • OPEN ACCESS

Properties and applications of the Kirkwood–Dirac distribution

To cite this article: David R M Arvidsson-Shukur *et al* 2024 *New J. Phys.* **26** 121201

View the article online for updates and enhancements.

New Journal of Physics
The open access journal at the forefront of physics

TOPICAL REVIEW

Properties and applications of the Kirkwood–Dirac distribution

David R M Arvidsson-Shukur[1,*], William F Braasch Jr[2], Stephan De Bièvre[3], Justin Dressel[4,5], Andrew N Jordan[4,5,6,7] ⓘ, Christopher Langrenez[3], Matteo Lostaglio[8], Jeff S Lundeen[9] ⓘ and Nicole Yunger Halpern[2,10] ⓘ

1   Hitachi Cambridge Laboratory, J. J. Thomson Avenue, Cambridge CB3 0HE, United Kingdom
2   Joint Center for Quantum Information and Computer Science, NIST and University of Maryland, College Park, MD 20742, United States of America
3   Univ. Lille, CNRS, Inria, UMR 8524, Laboratoire Paul Painlevé, F-59000 Lille, France
4   Institute for Quantum Studies, Chapman University, Orange, CA 92866, United States of America
5   Schmid College of Science and Technology, Chapman University, Orange, CA 92866, United States of America
6   The Kennedy Chair in Physics, Chapman University, Orange, CA 92866, United States of America
7   Department of Physics and Astronomy, University of Rochester, Rochester, NY 14627, United States of America
8   PsiQuantum, 700 Hansen Way, Palo Alto, CA 94304, United States of America
9   Department of Physics, Nexus for Quantum Technologies, University of Ottawa, Ottawa, Canada
10 Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742, United States of America
*   Author to whom any correspondence should be addressed.

E-mail: drma2@cam.ac.uk

Abstract

There are several mathematical formulations of quantum mechanics. The Schrödinger picture expresses quantum states in terms of wavefunctions over, e.g. position or momentum. Alternatively, phase-space formulations represent states with quasi-probability distributions over, e.g. position and momentum. A quasi-probability distribution resembles a probability distribution but may have negative and non-real entries. The most famous quasi-probability distribution, the Wigner function, has played a pivotal role in the development of a continuous-variable quantum theory that has clear analogues of position and momentum. However, the Wigner function is ill-suited for much modern quantum-information research, which is focused on finite-dimensional systems and general observables. Instead, recent years have seen the Kirkwood–Dirac (KD) distribution come to the forefront as a powerful quasi-probability distribution for analysing quantum mechanics. The KD distribution allows tools from statistics and probability theory to be applied to problems in quantum-information processing. A notable difference to the Wigner function is that the KD distribution can represent a quantum state in terms of arbitrary observables. This paper reviews the KD distribution, in three parts. First, we present definitions and basic properties of the KD distribution and its generalisations. Second, we summarise the KD distribution's extensive usage in the study or development of measurement disturbance; quantum metrology; weak values; direct measurements of quantum states; quantum thermodynamics; quantum scrambling and out-of-time-ordered correlators; and the foundations of quantum mechanics, including Leggett–Garg inequalities, the consistent-histories interpretation and contextuality. We emphasise connections between operational quantum advantages and negative or non-real KD quasi-probabilities. Third, we delve into the KD distribution's mathematical structure. We summarise the current knowledge regarding the geometry of KD-positive states (the states for which the KD distribution is a classical probability distribution), describe how to witness and quantify KD non-positivity, and outline relationships between KD non-positivity, coherence and observables' incompatibility.

Contents

1. Introduction

> *Negative energies and probabilities should not be considered as nonsense. They are well-defined concepts mathematically, like a negative sum of money.*
> – Paul Dirac in 1942 [1].

Phenomena such as non-commutation, coherence, and entanglement fundamentally distinguish quantum mechanics from classical physics. However, knowing exactly when an experiment lacks a classical analogue is notoriously difficult. If, and only if, joint probability distributions describe a system's preparation, manipulation and measurement, a corresponding experiment can be modelled classically [2–12].

Common strategies for pinpointing, understanding and developing quantum phenomena rely on quasi-probability distributions. A quasi-probability distribution enables the mathematical representation of a quantum state in terms of a joint distribution over the eigenvalues of possibly incompatible observables. Quasi-probability distributions satisfy some, but not all, of Kolmogorov's axioms of joint probability functions [13]. Some quasi-probability distributions can contain negative or even non-real elements. Such 'anomalous' values enable a probability-like description of quantum experiments and often herald non-classical phenomena.

In 1932, Wigner discovered his eponymous function, which has become the best-known quasi-probability distribution [4, 14–16]. Cousins of the Wigner function include the Glauber–Sudarshan [17, 18] and Husimi [19] representations. All three represent quantum states with distributions over the eigenvalues of conjugate observables that have continuous spectra, such as position and momentum, or an electromagnetic field's real and imaginary components. In many settings, the Wigner function assumes negative components when there is no classical description of the quantum state in a measurement scenario [20, 21]. Thus, the Wigner function has become a popular tool for studying quantum phenomena in continuous-variable systems—most notably in quantum optics [22].

Most modern quantum-information research, particularly quantum computing, concerns discrete-variable systems (e.g. qubit systems) and diverse observables (not merely position and momentum). The Wigner function is defined in terms of two fixed conjugate observables such as position and momentum, rather than arbitrary observables. (The discrete Wigner function [23] is defined in terms of two fixed maximally non-commuting observables.) When needing greater flexibility, one can employ the Kirkwood–Dirac (KD) distribution. Before we introduce the KD distribution and summarise its use cases, we briefly introduce its history.

In 1933, Kirkwood was inspired by the then-one-year-old Wigner function. He suggested a phase-space methodology for calculating thermodynamic partition functions [24]. Doing so, he introduced the quasi-probability distribution now known as the KD distribution. Shortly afterwards, in 1937, Terletsky independently introduced the real part of this distribution [25, 26]. In 1945, Dirac published an article that highlighted an 'analogy between classical and quantum mechanics' [5]. In this work, he independently rediscovered the KD distribution. Dirac argued that the only difference between classical and quantum mechanics stems from non-commutation [2, 3, 5]. Furthermore, he showed how the KD distribution can be used to calculate expectation values of functions of observables in a (quasi-)probabilistic framework [5].

After its original construction, the KD distribution received little attention. It was rediscovered again in 1968, by Rihaczek, within a classical signal-processing framework [27]. The KD distribution's real part, too, was rediscovered, by Margenau and Hill in 1961 [26, 28, 29]. Not until the recent development of quantum information theory did the KD distribution enjoy a substantial revival (e.g. [29–31]). It is now a common mathematical tool in quantum theory. The reason relates to Kirkwood's and Dirac's original motivations: The KD distribution provides a statistically-inspired framework for quantum calculations that are burdensome within the Schrödinger and Heisenberg formalisms.

Recent years have seen the KD distribution deployed to study or develop several areas of quantum mechanics. This Article's purpose is to review these past results. In quantum metrology (section 3), non-real KD quasi-probabilities are essential for accessing unknown information encoded in quantum states. Further, it is sometimes useful to distil metrological information from several quantum states into a few. If an underlying KD distribution contains negative values, the rate of this distillation can exceed classical limits (section 3.3). An example of such a technique is *weak-value amplification* (section 4), which can improve the signal-to-noise ratio (SNR) of measurements of weak-coupling parameters. Weak-value amplification achieves this improvement via the ability of KD negativity to boost a pre- and post-selected 'average' of an observable, such that the average lies outside the observable's spectrum. As an alternative to tomography, measurements of KD distributions (section 5) enable direct and/or efficient reconstructions of quantum states. In quantum thermodynamics (section 6), engines can operate in ways forbidden by classical theories when a certain KD distribution exhibits non-positivity. Furthermore, classical thermodynamics entails probability distributions over the possible amounts of work or entropy exchanged during a stochastic process. A KD distribution replaces such probability distributions in quantum thermodynamics. In quantum chaos (section 7), a generalised KD distribution can signal scrambling, the spreading of information about a local perturbation via many-body entanglement. Finally, the KD distribution surfaces also in foundational settings (section 8). 'Temporal Bell inequalities', or Leggett–Garg inequalities, can be violated only when underlying KD distributions are non-positive (section 8.1). Moreover, in the consistent-histories interpretation of quantum mechanics, cross-terms in a KD distribution govern whether a classical phase-space history can be ascribed to a quantum state (section 8.2). A rigorous notion of non-classicality is quantum contextuality. Several proofs of experimental contextuality rely on non-positive KD distributions directly or indirectly (section 8.3).

Thanks to this growing interest, it has become crucial to understand the mathematical properties of the KD distribution. Sufficient and necessary requirements for the Wigner function to be negative have been known since 1976 [7]. However, such requirements for pure states' KD distributions were mapped out only recently [32–34]. The extension to mixed states is not yet fully understood [35, 36]. Section 9 reviews mathematical results about positive and non-positive KD distributions. We summarise the current understanding of the geometry of KD-positive states. Also, we outline how to witness and quantify KD non-positivity and coherence.

Different readers will benefit from using this review differently. Reading the entire paper, one gains a holistic view of the KD distribution. Reading individual sections in a modular fashion, one can target the distribution's most relevant applications and properties. We introduce the KD distribution and outline its basic properties before reviewing use cases.

2. Definition and basic properties

In this section, we define the KD distribution. The KD distribution is a versatile and adaptable object that can assume multiple forms. For simplicity, we introduce, first, the most common form of the KD distribution. Then, we describe how to construct a general KD distribution. The KD distribution is applicable to discrete and continuous-variable systems. However, this review centres on discrete systems. Indeed, the KD distribution has been applied most to discrete systems. Nevertheless, section 2.3 covers continuous-variable KD distributions, as needed for our review of wavefunction measurements (section 5). Section 9 introduces further properties of the KD distribution.

2.1. Quasi-probability representations of quantum mechanics

Here, we briefly overview quasi-probability representations of quantum mechanics. (See [37] and references therein for an extensive discussion.) Then, we define the KD representation of quantum mechanics and the KD quasi-probability distribution. Less mathematically inclined readers may skip to the next subsection.

To introduce quasi-probability representations of quantum mechanics, we must first provide some definitions. Denote by $\mathcal{H}$ a d-dimensional Hilbert space, by $\hat{\rho}$ an arbitrary density operator defined on $\mathcal{H}$ and by $\hat{C}$ an arbitrary linear operator defined on $\mathcal{H}$. Additionally, let us choose a set Λ. What Λ consists of depends on the specific quasi-probability representation, as illustrated across this review. Examples include tuples of incompatible observables' eigenvalues (section 2.2), hidden variables (section 8.3) and position–momentum phase space (section 2.3 and [4]). (For notational simplicity, we assume Λ to be finite from now on.) We associate to each density operator $\hat{\rho}$ a complex-valued function $\lambda \to Q_\lambda(\hat{\rho})$ on Λ, and to each linear operator $\hat{C}$ a function $\lambda \to T_\lambda(\hat{C})$ on Λ, such that

$$\mathrm{Tr}\left(\hat{C}^\dagger \hat{\rho}\right) = \sum_\lambda T_\lambda^*\left(\hat{C}\right) Q_\lambda\left(\hat{\rho}\right). \tag{1}$$

The function $Q(\hat{\rho}): \lambda \to Q_\lambda(\hat{\rho})$ is a *quasi-probability distribution* (or quasi-probability density) for $\hat{\rho}$, and $T(\hat{C}): \lambda \to T_\lambda(\hat{C})$ a *symbol* for $\hat{C}$. Together, the maps $\hat{\rho} \to Q(\hat{\rho})$ and $\hat{C} \to T(\hat{C})$ form a quasi-probability representation of quantum mechanics on $\mathcal{H}$.

Depending on the physical setup, quasi-probability representations may obey extra requirements. One typically assumes that $\hat{\rho} \to Q(\hat{\rho})$ is convex-linear and that $\hat{C} \to T(\hat{C})$ is linear. Additionally, the identity operator's symbol is often a constant, equal to one, in λ:

$$T_\lambda\left(\hat{\mathbb{1}}\right) = 1, \ \ \forall \lambda \in \Lambda. \tag{2}$$

This equality, with equation (1), implies the quasiprobability distribution's normalisation:

$$\mathrm{Tr}\left(\hat{\rho}\right) = \sum_\lambda Q_\lambda\left(\hat{\rho}\right) = 1, \forall \hat{\rho}. \tag{3}$$

An occasional requirement is the reality of every observable's symbol:

$$T_\lambda\left(\hat{C}^\dagger\right) = T_\lambda^*\left(\hat{C}\right), \ \ \forall \lambda \in \Lambda. \tag{4}$$

Several quasi-probability distributions satisfy equation (4), as explained in [37]. KD distributions (defined below) can be non-real. Thus, they fall outside the framework in [37], which concerns real-valued quasi-probabilities. Nevertheless, much of the framework in [37] can be straightforwardly extended to encompass also complex-valued quasi-probability distributions.

Equation (1) forms the quasi-probability representation's central ingredient. Using it, one can calculate quantum expectation values by calculating averages of random variables $T(\hat{C})$ with respect to complex quasi-probability measures $Q(\hat{\rho})$. Thus, the quasi-probability distribution's role in expectation-value calculations parallels a positive probability measure's role in classical mechanics. Here, however, the measures need not be positive; whence the terminology 'quasi-probability'. Moreover, quasi-probability representations can be straightforwardly enriched to represent operations and evolutions [38, 39].

We now define a quantum state's KD distribution and show how it fits into the above-outlined framework. We consider two orthonormal bases in $\mathcal{H}$: $\{|a_i\rangle\}$ and $\{|b_j\rangle\}$, wherein $i, j = 1, \ldots d$. Without loss

of generality, we can view $\{|a_i\rangle\}$ and $\{|b_j\rangle\}$ as eigenbases of Hermitian observables $\hat{A}$ and $\hat{B}$. For now, we assume that these observables are non-degenerate. They, and the identity operator, eigendecompose as

$$\hat{A} = \sum_i a_i |a_i\rangle\langle a_i|, \quad \hat{B} = \sum_j b_j |b_j\rangle\langle b_j| \quad \text{and} \quad \hat{\mathbb{1}} = \sum_i |a_i\rangle\langle a_i| = \sum_j |b_j\rangle\langle b_j|. \tag{5}$$

We write discrete sums here for simplicity. However, one can generalise the equations to continuous bases and/or infinite-dimensional Hilbert spaces: one replaces the summations with integrals over an appropriate measure (see section 2.3). We will suppose that $\langle a_i|b_j\rangle \neq 0$ for all i, j. One can then express an arbitrary operator $\hat{C}$ in terms of both bases:

$$\hat{C} = \left(\sum_i |a_i\rangle\langle a_i|\right) \hat{C} \left(\sum_j |b_j\rangle\langle b_j|\right) = \sum_{i,j} \langle a_i|\hat{C}|b_j\rangle |a_i\rangle\langle b_j| = \sum_{i,j} \langle b_j|a_i\rangle\langle a_i|\hat{C}|b_j\rangle \frac{|a_i\rangle\langle b_j|}{\langle b_j|a_i\rangle} \equiv \sum_{i,j} C(a_i, b_j) \hat{G}_{a_i, b_j}. \tag{6}$$

The coefficients $C(a_i, b_j)$ and normalised outer products $\hat{G}_{a_i, b_j}$ are

$$C(a_i, b_j) \equiv \langle b_j|a_i\rangle\langle a_i|\hat{C}|b_j\rangle \quad \text{and} \quad \hat{G}_{a_i, b_j} \equiv \frac{|a_i\rangle\langle b_j|}{\langle b_j|a_i\rangle}. \tag{7}$$

The set $\{\hat{G}_{a_i, b_j} \mid 1 \leqslant i, j \leqslant d\}$ forms a basis for the space of linear operators defined on $\mathcal{H}$. However, the $\hat{G}_{a_i, b_j}$ are generally not orthogonal projectors or self-adjoint. Nevertheless, $\{\hat{G}_{a_i, b_j}\}$ forms an orthogonal basis for the space of operators defined on $\mathcal{H}$, with respect to the Hilbert-Schmidt inner product. In terms of this basis, $\hat{C}$ has the component representation $C(a_i, b_j)$.

Suppose that the operator $\hat{C}$ is a quantum state $\hat{\rho}$. The components $C(a_i, b_j)$ form the *standard* KD quasi-probability distribution of $\hat{\rho}$. We denote it by $Q(\hat{\rho})$:

$$Q_{a_i, b_j}(\hat{\rho}) \equiv \langle b_j|a_i\rangle\langle a_i|\hat{\rho}|b_j\rangle. \tag{8}$$

The KD symbol of an arbitrary linear operator $\hat{C}$ is

$$T_{a_i, b_j}(\hat{C}) \equiv \frac{\langle a_i|\hat{C}|b_j\rangle}{\langle a_i|b_j\rangle}. \tag{9}$$

If $\hat{C}$ is an observable, the $T_{a_i, b_j}(\hat{C})$ are weak values (see section 4). Using equations (8) and (9), one can verify equation (1): $\text{Tr}(\hat{C}^\dagger \hat{\rho}) = \sum_{a_i, b_j} T_{a_i, b_j}(\hat{C})^* Q_{a_i, b_j}(\hat{\rho})$. Consequently, (Q, T) forms a quasi-probability representation of quantum mechanics over the set $\Lambda = \{(a_i, b_j) \mid 1 \leqslant i, j \leqslant d\}$. We call this representation the *standard KD representation*, which depends on $\hat{A}$ and $\hat{B}$. This representation satisfies equations (2) and (3) but not equation (4).

We now detail further the structure of the standard KD representation of quantum mechanics. Define the projector $\hat{\Pi}_{a_i} = |a_i\rangle\langle a_i|$ and the projector product $\hat{F}_{a_i, b_j} = \hat{\Pi}_{a_i}\hat{\Pi}_{b_j}$. The set $\{\hat{F}_{a_i, b_j}\}$ forms an orthogonal basis for the space of operators defined on $\mathcal{H}$: $\text{Tr}(\hat{F}^\dagger_{a_{i'}, b_{j'}} \hat{F}_{a_i, b_j}) = |\langle a_i|b_j\rangle|^2 \delta_{i, i'}\delta_{j, j'}$. This basis is a particular instance of a *frame*. Dual to $\{\hat{F}_{a_i, b_j}\}$ is the *dual basis* $\{\hat{G}_{a_i, b_j}\}$, since $\text{Tr}(\hat{G}^\dagger_{a_i, b_j}\hat{F}_{a_{i'}, b_{j'}}) = \delta_{i, i'}\delta_{j, j'}$. The dual basis' elements have the form $\hat{G}_{a_i, b_j} = \hat{\Pi}_{a_i}\hat{\Pi}_{b_j}/|\langle a_i|b_j\rangle|^2$. Concise expressions for $Q(\hat{\rho})$ and $T(\hat{C})$ follow:

$$Q_{a_i, b_j}(\hat{\rho}) = \text{Tr}\left(\hat{F}^\dagger_{a_i, b_j}\hat{\rho}\right), \quad \text{and} \quad T_{a_i, b_j}(\hat{C}) = \text{Tr}\left(\hat{G}^\dagger_{a_i, b_j}\hat{C}\right). \tag{10}$$

Now, we can express states $\hat{\rho}$ in terms of quasi-probabilities $Q(\hat{\rho})$ and observables $\hat{C}$ in terms of symbols $T(\hat{C})$. The corresponding *reconstruction formulae* are

$$\hat{\rho} = \sum Q_{a_i, b_j}(\hat{\rho})\hat{G}_{a_i, b_j}, \quad \text{and} \quad \hat{C} = \sum_{a_i, b_j} T_{a_i, b_j}(\hat{C})\hat{F}_{a_i, b_j}. \tag{11}$$

Ferrie [37] details the construction of quasi-probability representations using frames and their duals. In no quasi-probability representation are all $Q(\hat{\rho})$ positive measures and all $T_{a_i, b_j}(\hat{C}) \geqslant 0$ for all i, j and all $\hat{C} \geqslant 0$. For a more precise statement, see, e.g. theorem 4 of [37].

In [5], Dirac aimed to define a function of non-commuting observables $\hat{A}$ and $\hat{B}$. The 'ordering problem' of quantum mechanics motivated Dirac: The position operator $\hat{x}$ and momentum operator $\hat{p}$ do not commute. Therefore, how can one associate an operator $\hat{F}(\hat{x}, \hat{p})$ to a function $f(x, p)$ of the classical position x and momentum p? That is, how does one 'quantise' functions of classical observables? If f is a polynomial,

one can, e.g. left-order the variables: One would place all powers of $\hat{x}$ leftward of all powers of $\hat{p}$. An alternative is right-ordering. Wigner proposed a symmetric ordering, embodied in the Wigner function [4]. Dirac introduced left-ordering and right-ordering of general non-commuting operators $\hat{A}$ and $\hat{B}$. We can rephrase Dirac's proposal as follows. Consider non-commuting observables $\hat{A}$ and $\hat{B}$, as well as a function $h(a_i, b_j)$ of the eigenvalue pairs $(a_i, b_j) \in \mathbb{R}^2$. Define the operator $\hat{h}(\hat{A}, \hat{B})$ as

$$\hat{h}\left(\hat{A}, \hat{B}\right) \equiv \sum_{a_i, b_j} h\left(a_i, b_j\right) \hat{F}_{a_i, b_j}. \tag{12}$$

This definition resembles the second equation in equations (11). In Dirac's language, every operator $\hat{C}$ is a function $\hat{h}(\hat{A}, \hat{B})$ of $\hat{A}$ and $\hat{B}$. The symbol $T(\hat{C})$ plays the role of h:

$$\hat{C} = \hat{h}\left(\hat{A}, \hat{B}\right), \quad \text{wherein} \quad h\left(a_i, b_j\right) = T_{a_i, b_j}\left(\hat{C}\right). \tag{13}$$

When $h(a_i, b_j) = f(a_i)g(b_j)$, $\hat{h}(\hat{A}, \hat{B}) = \hat{f}(\hat{A})\hat{g}(\hat{B})$. In other words, Dirac's prescription amounts to an ordering '$\hat{A}$ to the left of $\hat{B}$'. One can order the operators oppositely instead: one replaces $\hat{F}_{a_i, b_j}$ by its adjoint, $\hat{F}^\dagger_{a_i, b_j} = \hat{\Pi}_{b_j}\hat{\Pi}_{a_i}$.

2.2. The standard KD distribution

The previous subsection introduced the standard KD distribution $Q(\hat{\rho})$ for a quantum state $\hat{\rho}$ with respect to bases $\{|a_i\rangle\}$ and $\{|b_j\rangle\}$. We can cast this distribution as a $d \times d$-dimensional matrix with entries

$$Q_{i,j}(\hat{\rho}) \equiv \langle b_j | a_i \rangle \langle a_i | \hat{\rho} | b_j \rangle. \tag{14}$$

In the left-hand side, we have replaced a_i with i, and b_j with j, for notational simplicity. Equation (14) does not depend upon whether $\{|a_i\rangle\}$ and $\{|b_j\rangle\}$ are discrete or continuous. As mentioned above, this review focuses on finite dimensions; the continuous KD distribution is introduced in section 2.3 and used only in section 5. Suppose that $\langle b_j | a_i \rangle \neq 0$ for all i, j. By the reconstruction formulae equations (11), $Q(\hat{\rho})$ enables an informationally complete description of $\hat{\rho}$:

$$\hat{\rho} = \sum_{i,j} \hat{G}_{a_i, b_j} Q_{i,j}(\hat{\rho}). \tag{15}$$

As explained above, $\{\hat{G}_{a_i, b_j}\} = \{|a_i\rangle\langle b_j| / \langle b_j | a_i \rangle\}$ forms a basis. Now, suppose that some $\langle b_j | a_i \rangle = 0$. The KD distribution can still convey useful, albeit partial, information about $\hat{\rho}$. Examples surface in weak-value experiments (section 4), where one need not necessarily apply an entire KD distribution.

The KD distribution satisfies several of Kolmogorov's axioms [13] for joint probability distributions:

$$\sum_{i,j} Q_{i,j}(\hat{\rho}) = 1, \quad \sum_j Q_{i,j}(\hat{\rho}) = \langle a_i | \hat{\rho} | a_i \rangle \quad \text{and} \quad \sum_i Q_{i,j}(\hat{\rho}) = \langle b_j | \hat{\rho} | b_j \rangle. \tag{16}$$

Nevertheless, $Q(\hat{\rho})$ is not a joint probability distribution: Although $|Q_{i,j}(\hat{\rho})| \in [0,1]$, $Q_{i,j}(\hat{\rho})$ can assume negative or non-real values. Still, the marginals of $Q(\hat{\rho})$ are probability distributions. Also, they reproduce the Born-rule probabilities associated to $\hat{A}$ and $\hat{B}$. This simple reproduction is not a requirement for quasi-probability representations generally, but a special and important feature of the KD distributions. The Wigner function, too, reproduces the Born-rule probabilities via marginalisation [4]. To reproduce the probabilities with the Husimi function or Glauber-Sudarshan distribution, one must undertake a more extended calculation [17–19]. Furthermore, the KD representation of quantum mechanics does not obey equation (4). Consequently, self-adjoint operators do not necessarily have real KD symbols.

The negative and non-real values of $Q_{i,j}(\hat{\rho})$ are sometimes called 'non-classical'. What is considered non-classical differs from setting to setting (and author to author). Therefore, we refer to negative or non-real values simply as *non-positive*. We call $Q(\hat{\rho})$ *positive* if all its entries are positive or zero. The most common measure of KD non-positivity is [40]

$$\mathcal{N}(Q(\hat{\rho})) = \sum_{i,j} \left| Q_{i,j}(\hat{\rho}) \right|. \tag{17}$$

$\mathcal{N}(Q(\hat{\rho})) = 1$ if, and only if, the KD distribution is a classical joint probability distribution. Section 9 outlines further properties of $\mathcal{N}(Q(\hat{\rho}))$.

2.3. Continuous-variable KD distribution

Throughout this article, we denote discrete-variable projectors by $\hat{\Pi}$ and continuous-variable projectors by $\hat{\pi}$. We denote by $P(a)$ a probability distribution over a discrete variable a and by $\mathbb{P}(x)$ a probability density over a continuous variable x. For continuous-variable quantum states, the KD distribution is typically constructed in terms of the phase space's canonical conjugate observables: position,

$$\hat{x} = \int x\hat{\pi}_x \mathrm{d}x \equiv \int x|x\rangle\langle x|\mathrm{d}x, \tag{18}$$

and momentum,

$$\hat{p} = \int p\hat{\pi}_p \mathrm{d}p \equiv \int p|p\rangle\langle p|\mathrm{d}p. \tag{19}$$

The operators satisfy $[\hat{x},\hat{p}] = i\hbar\hat{\mathbb{1}}$. The corresponding basis elements of equation (7) (with $\hat{A} = \hat{x}$ and $\hat{B} = \hat{p}$) have the form

$$\hat{G}(x,p) = \frac{|x\rangle\langle p|}{\langle p|x\rangle} = \sqrt{2\pi\hbar}\, \mathrm{e}^{ipx/\hbar}|x\rangle\langle p| = 2\pi\hbar\,\hat{\pi}_x\hat{\pi}_p\,. \tag{20}$$

In terms of them, we can expand the density operator:

$$\hat{\rho} = \iint Q(x,p;\hat{\rho})\,\hat{G}(x,p)\,\mathrm{d}x\,\mathrm{d}p. \tag{21}$$

$Q(x,p;\hat{\rho})$ denotes the continuous-variable KD distribution:

$$Q(x,p;\hat{\rho}) \equiv \langle p|x\rangle\langle x|\hat{\rho}|p\rangle = \mathrm{Tr}\left(\hat{\pi}_p\hat{\pi}_x\hat{\rho}\right). \tag{22}$$

We interpret this formula in section 5. As for discrete KD distributions, any marginalisation (an integration over momentum or position) completes the Fourier transform. The position probability density, $\mathbb{P}(x) = \mathrm{Tr}\left(\hat{\pi}_x\hat{\rho}\right)$, or the momentum probability density, $\mathbb{P}(p) = \mathrm{Tr}\left(\hat{\pi}_p\hat{\rho}\right)$, results.

Consequently, a pure state $|\psi\rangle$ has a KD distribution of the form

$$Q(x,p;\psi) \equiv \langle p|x\rangle\langle x|\psi\rangle\langle\psi|p\rangle = \frac{\mathrm{e}^{-ipx/\hbar}}{\sqrt{2\pi\hbar}}\psi(x)\widetilde{\psi}^*(p). \tag{23}$$

$Q(x,p;\psi) \equiv \langle p|x\rangle\langle x|\psi\rangle\langle\psi|p\rangle$ combines the position wavefunction $\psi(x) = \langle x|\psi\rangle$ and the momentum wavefunction $\widetilde{\psi}(p) = \langle p|\psi\rangle$ with the Fourier kernel that connects the two bases. In summary, despite subtle differences, the continuous KD distribution straightforwardly extends the discrete KD distribution.

We can now specify how to quantise functions of non-commuting observables. Equation (12) becomes

$$\hat{h}(\hat{x},\hat{p}) = \int h(x,p)\,\hat{\pi}_x\hat{\pi}_p\,\mathrm{d}x\,\mathrm{d}p. \tag{24}$$

Equation (24) constitutes a (Kohn Nirenberg) quantisation [41] in which a polynomial in p, with x-dependent coefficients, is transformed into a differential operator. Interchanging the $\hat{\pi}_x$ and $\hat{\pi}_p$ yields the opposite ordering.

2.4. Quasi-probabilistic Bayesian update

In classical statistics, Bayes' theorem dictates how one updates a joint probability distribution upon acquiring new information. The KD distribution obeys a quasi-probabilistic version of Bayes' theorem [42–45]. This version updates the distribution when one learns a restriction on i or j. We define the conditional quasi-probabilities $\tilde{Q}_{i|j}(\hat{\rho})$ and $\tilde{Q}_{j|i}(\hat{\rho})$ by

$$\tilde{Q}_{i|j}(\hat{\rho}) \equiv \frac{Q_{i,j}(\hat{\rho})}{P(b_j|\hat{\rho})} \quad \text{and} \quad \tilde{Q}_{j|i}(\hat{\rho}) \equiv \frac{Q_{i,j}(\hat{\rho})}{P(a_i|\hat{\rho})}. \tag{25}$$

We interpret $\tilde{Q}_{i|j^\star}(\hat{\rho})$ as the KD distribution $Q_{i,j}(\hat{\rho})$ conditioned on $j = j^\star$. We have defined $P(b_j|\hat{\rho}) = \sum_i Q_{i,j}(\hat{\rho}) = \langle b_j|\hat{\rho}|b_j\rangle$ as the probability of obtaining the outcome b_j upon measuring the state $\hat{\rho}$ in the $\{|b_j\rangle\}$ basis. We have defined $P(a_i|\hat{\rho})$ analogously. If $Q(\hat{\rho})$ is positive, then $|\tilde{Q}_{i|j}(\hat{\rho})| \in [0,1]$ for all i,j, as expected from classical probability theory. However, a KD distribution can assume negative values. Therefore, the denominator in equation (25) can have a lesser magnitude than the numerator. Consequently, $|\tilde{Q}_{i|j}(\hat{\rho})| \in [0,\infty)$, in general. As we review below, such anomalous Bayesian updates explain several

non-classical advantages in quantum-information processing. Moreover, the conditional-KD-distribution formulae suffice for reformulating a quantum evolution as a quasi-probability update for classical (eigen)values [44, 46]. Bayesian evolution of the KD distribution was experimentally demonstrated via direct measurement [47], a method described in section 5.

More generally, one can condition $Q_{i,j}(\hat{\rho})$ such that j assumes a value from a subset $\mathcal{F}$ [48, 49]:

$$\tilde{Q}_{i,j|j\in\mathcal{F}}(\hat{\rho}) = \frac{Q_{i,j}(\hat{\rho})}{\sum_{i'}\sum_{j'\in\mathcal{F}}Q_{i',j'}(\hat{\rho})}, \tag{26}$$

if j is in $\mathcal{F}$. Otherwise, $\tilde{Q}_{i,j|j\in\mathcal{F}}(\hat{\rho}) = 0$. The denominator is the probability that, if $\hat{\rho}$ is prepared and $\{|b_j\rangle\}$ is measured, the outcome j will be in $\mathcal{F}$.

2.5. Generalisations of the KD distribution

The distribution in equation (14) is, from a fundamental perspective, the most studied and mathematically understood KD distribution. Nevertheless, generalisations of the standard KD distribution have been used frequently in applications. The first generalisation [5, 24] allows one to decompose a quantum state with respect to more than two sets of measurement operators. The second generalisation [32, 45, 50] allows for these sets to be not only rank-1 projectors, but general positive-operator-valued measures.

Consider k non-degenerate observables $\hat{A}^{(l)} = \sum_{i_l=1}^{d} a_{i_l}^{(l)} |a_{i_l}^{(l)}\rangle\langle a_{i_l}^{(l)}|$, where $l = 1, 2, \ldots, k$. We denote the corresponding eigenbases by $\{|a_{i_l}^{(l)}\rangle\}$. The k-extended KD distribution [45] is

$$Q_{i_1,\ldots,i_k}(\hat{\rho}) = \langle a_{i_k}^{(k)}|a_{i_{k-1}}^{(k-1)}\rangle\langle a_{i_{k-1}}^{(k-1)}|a_{i_{k-2}}^{(k-2)}\rangle\cdots\langle a_{i_1}^{(1)}|\hat{\rho}|a_{i_k}^{(k)}\rangle. \tag{27}$$

Equation (27) shows that, to write an extended KD distribution, we must choose how to order the observables. The physical problem of interest should determine the ordering. For examples, see the extended KD distributions constructed in section 3.3 (equation (38)) and section 7 (equation (70)).

The second generalisation of the KD distribution involves measurement operators that are not necessarily rank-1 projectors, but form general positive-operator-valued measures. A positive-operator-valued measure is a set $\{\hat{M}_i\}$ of positive-semidefinite operators $\hat{M}_i \geqslant 0$ that are normalised: $\sum_i \hat{M}_i = \hat{\mathbb{1}}$ [51]. Positive-operator-valued measures represent arbitrary quantum measurements, not only projective measurements. In terms of k positive-operator-valued measures $\mathcal{M}^{(l)} = \{\hat{M}_{i_l}^{(l)}\}$, where $l = 1, 2, \ldots, k$, the measurement-generalised KD distribution is

$$Q_{i_1,\ldots,i_k}(\hat{\rho}) = \mathrm{Tr}\left(\hat{M}_{i_k}^{(k)}\hat{M}_{i_{k-1}}^{(k-1)}\cdots\hat{M}_{i_1}^{(1)}\hat{\rho}\right). \tag{28}$$

Equation (28) shows the most general form of a KD distribution (yet defined). The extended and measurement-generalised KD distributions satisfy trivially extended versions of the properties listed in equations (16) and (25). The non-positivity of $\{Q_{i_1,\ldots,i_k}(\hat{\rho})\}$ is often quantified by $\mathcal{N}(Q(\hat{\rho})) = \sum_{i_1,\ldots,i_k}|Q_{i_1,\ldots,i_k}(\hat{\rho})|$, as outlined in section 9.3. In this review, we denote by $Q(\hat{\rho})$ standard, extended and measurement-generalised KD distributions. The context and the number of indices will specify a distribution further.

2.6. Optimisation with the KD distribution

A common use of the KD distribution is the evaluation and optimisation of operational formulae. Consider some physical formula $F(\hat{\rho})$ of interest. One might recast it in terms of a KD distribution: $\tilde{F}(Q(\hat{\rho}))$. If so, one can optimise $\tilde{F}(Q(\hat{\rho}))$ with respect to a classical probability distribution or a general KD distribution:

$$F^{\mathrm{p}} \equiv \underset{Q_{i,j}\in[0,1]}{\mathrm{opt}}\left\{\tilde{F}(Q)\right\}, \qquad \text{or} \qquad F^{\mathrm{np}} \equiv \underset{Q(\hat{\rho})}{\mathrm{opt}}\left\{\tilde{F}(Q(\hat{\rho}))\right\}, \tag{29}$$

The opt could entail a maximisation, a minimisation or some other optimisation procedure. If F^{np} differs from F^{p}, then non-commutation (in the form of KD non-positivity) can break the bound on the value achievable in scenarios describable with classical probability distributions. Moreover, the optimised KD distribution,

$$Q^\star(\hat{\rho}) = \arg\underset{Q(\hat{\rho})}{\mathrm{opt}}\left\{\tilde{F}(Q(\hat{\rho}))\right\}, \tag{30}$$

guides the construction of an optimal experiment. In the next sections, we will see several examples.
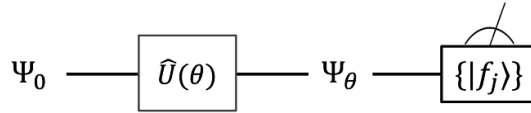
Figure 1. Standard quantum metrology. The goal is to use a quantum system as a probe to learn an unknown parameter of a unitary. Initially, the probe state is $|\Psi_0\rangle$. The probe undergoes a unitary $\hat{U}(\theta)$, which encodes the unknown parameter θ in the probe state: $|\Psi_0\rangle \rightarrow |\Psi_\theta\rangle = \hat{U}(\theta)|\Psi_0\rangle$. Measuring the state in a suitable basis yields information about θ.

3. The KD distribution and quantum metrology

Metrology is the science of measuring or estimating unknown physical parameters. When one estimates parameters that characterise quantum processes, it is natural to use quantum systems as probes. Measurements of unknown parameters, using quantum systems, fall under the domain of quantum metrology. In quantum metrology, non-classical phenomena, such as coherence, entanglement and non-commutation, can boost estimation abilities beyond classical bounds [48, 52–60]. Below, we review how non-real entries in KD distributions play a fundamental role in quantum metrology. We also show that one can break classical bounds on metrological quantum-information distillation only in the presence of KD non-positivity [48]. For simplicity, we focus on pure states, Stone's-encoded unitaries (explained below) and single-parameter metrology. However, generalisations extend the results beyond these restrictions [49, 50, 61]. Further applications lie beyond the scope of this review: for example, KD negativity can quantify resources in interaction-free measurements [62].

3.1. Measurement disturbance

Before we review the KD distribution's connection to quantum metrology, we describe how imaginary KD quasi-probabilities encode the disturbance of a quantum state. Let θ denote a real parameter that we wish to measure. Suppose that a unitary $\hat{U}(\theta)$ obeys Stone's theorem [63]: $\hat{U}(\theta) = e^{-i\hat{A}\theta}$, wherein $\hat{A} = \sum_i a_i |a_i\rangle\langle a_i|$ denotes a Hermitian generator. We assume that the eigenvalues a_i are non-degenerate. Consider evolving a state $|\Psi_0\rangle$ under $\hat{U}(\theta)$ to the output state $|\Psi_\theta\rangle = \hat{U}(\theta)|\Psi_0\rangle$. Finally, consider measuring some basis $\{|f_j\rangle\}$. (See figure 1.) This process yields a probability distribution $\{P(f_j|\Psi_\theta) = |\langle f_j|\Psi_\theta\rangle|^2\}$.

A natural question to ask is *how much do changes in θ disturb the measurement-outcome probabilities?* The question invites us to differentiate:

$$
\begin{aligned}
\partial_\theta P\left(f_j|\Psi_\theta\right) &= \partial_\theta |\langle f_j|\Psi_\theta\rangle|^2 \\
&= -i\langle f_j|\hat{A}|\Psi_\theta\rangle\langle\Psi_\theta|f_j\rangle + i\langle f_j|\Psi_\theta\rangle\langle\Psi_\theta|\hat{A}|f_j\rangle \\
&= 2\mathrm{Im}\left(\langle f_j|\hat{A}|\Psi_\theta\rangle\langle\Psi_\theta|f_j\rangle\right) \\
&= \sum_i 2a_i\mathrm{Im}\left(\langle f_j|\Psi_\theta\rangle\langle\Psi_\theta|a_i\rangle\langle a_i|f_j\rangle\right) \\
&= \sum_i 2a_i\mathrm{Im}\left(Q_{i,j}(\Psi_\theta)\right).
\end{aligned}
\tag{31}
$$

We have defined the KD quasi-probability $Q_{i,j}(\Psi_\theta) \equiv \langle f_j|\Psi_\theta\rangle\langle\Psi_\theta|a_i\rangle\langle a_i|f_j\rangle$. A generator's ability to disturb a quantum state is related to a KD distribution's imaginary part. For further results on this topic, see [64]. This reference describes how imaginary weak values (which stem from imaginary KD distributions, as outlined in section 4) signal a von Neumann measurement's disturbance of a quantum state. Moreover, [65, 66] discuss connections between the imaginary part of a KD distribution and a quantum state's asymmetry.

3.2. Standard quantum metrology

Again, we consider a unitary evolution $\hat{U}(\theta) = e^{-i\hat{A}\theta}$, where θ is now an unknown parameter. To estimate θ, one can prepare a probe state $|\Psi_0\rangle$, evolve it under $\hat{U}(\theta)$ and measure the final state, $|\Psi_\theta\rangle = \hat{U}(\theta)|\Psi_0\rangle$. Again, the measurements are of some basis $\{|f_j\rangle\}$. Repeating this process across many trials provides metrologically useful statistics (see figure 1). The measurement outcomes obey a probability distribution $\{P(f_j|\Psi_\theta) = |\langle f_j|\Psi_\theta\rangle|^2\}$.

Sampling from this distribution, one observes the probabilities $\{P_o(f_j|\Psi_\theta)\}$. An estimator θ_e is a function that maps the sample space (the set of possible outcomes that may be outputted during the experiment) to a sample estimate (the θ estimate inferred from the observations). In other words, given observed data, one can construct an estimate $\theta_e(\{P_o(f_j|\Psi_\theta)\})$ of the unknown parameter θ.
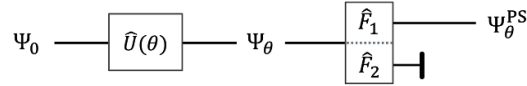
Figure 2. Post-selected quantum metrology. The goal is to distil the metrological information from many quantum systems into a few. Information about an unknown parameter θ is encoded in quantum states, as in figure 1. The information-carrying states then undergo a two-outcome positive-operator-valued measure $\{\hat{F}_1 \equiv \hat{F}, \hat{F}_2 \equiv \hat{\mathbb{1}} - \hat{F}\}$, which acts as a filter. Probes that pass the post-selection, yielding the $\hat{F}_1$ outcome, are retained (post-selected). Probes that do not are discarded. If $\hat{F}$ is chosen appropriately, the quantum Fisher information of a post-selected state $|\Psi_\theta^{\mathrm{PS}}\rangle$ can be significantly higher than that of $|\Psi_\theta\rangle$. The post-selection probability ensures that, on average, the post-selective filter creates no information. Moreover, consider operating on n states with the post-selective filter, and suppose that m states pass the post-selection. In certain cases, the amount of information retained in the post-selected states can nearly equal the total amount of information available initially [49].

The precision of any unbiased estimator, classical or quantum, is lower-bounded by the Cramér–Rao inequality:

$$\mathrm{Var}\left(\theta_{\mathrm{e}}\right) \geqslant \frac{1}{NI(\theta)}. \tag{32}$$

N denotes the number of samples (measurements) drawn from the probability distribution, and $I(\theta)$ denotes the Fisher information [67, 68]. The Fisher information of a probability distribution $\{P(f_j|\Psi_\theta)\}$ is

$$I(\theta) = \sum_j \frac{\left[\partial_\theta P\left(f_j|\Psi_\theta\right)\right]^2}{P\left(f_j|\Psi_\theta\right)}. \tag{33}$$

Inequality (32) saturates for large N and reasonable estimators. Thus, one can improve an estimate in two ways. First, one can increase the number N of measurements. Second, one can choose the initial state $|\Psi_0\rangle$ and the final-measurement basis $\{|f_j\rangle\}$ such that the Fisher information is large.

For any metrological estimation to be possible, the Fisher information about θ, obtained from measuring $|\Psi_\theta\rangle$, must be non-zero. Such a non-zero Fisher information is directly connected to the KD distribution. In equation (33), the numerator is the square of the derivative of the outcome probability (31). Since $\tilde{Q}_{i|j}(\Psi_\theta) \equiv \frac{Q_{i,j}(\Psi_\theta)}{P(f_j|\Psi_\theta)}$ (equation (25)), we can re-express equation (33) as

$$I(\theta) = 4 \sum_j \frac{\left[\sum_i a_i \mathrm{Im}\left(Q_{i,j}(\Psi_\theta)\right)\right]^2}{P\left(f_j|\Psi_\theta\right)} = 4 \sum_j P\left(f_j|\Psi_\theta\right) \left[\sum_i a_i \mathrm{Im}\left(\tilde{Q}_{i|j}(\Psi_\theta)\right)\right]^2. \tag{34}$$

Thus, a non-zero Fisher information (obtained by measuring $|\Psi_\theta\rangle$) requires *non-real* components in a conditional KD distribution $\tilde{Q}_{i|j}(\Psi_\theta)$. Consequently, $\{|f_j\rangle\}$ cannot equal $\{|a_i\rangle\}$, if an experiment is to extract information about θ: Such equality would lead to a positive KD distribution. In other words, if the measurement basis equals the basis of $\hat{A}$, then the final probabilities are unchanged by the unitary evolution and hence do not depend on θ.

Hofmann [69] discusses further links among the Fisher information, KD distributions and weak values.

3.3. Post-selected quantum metrology

In some quantum experiments, probe systems are more easily prepared than measured. This is the case, for example, in many optics experiments, where heralded single photons can be created significantly faster than they can be measured [50, 70]. The difficulty in measuring single photons stems from particle detectors' saturation-intensity limit, or dead time: after a detector detects a particle, it experiences a time lag (dead time) until it can detect another particle. Moreover, there is a maximum energy intensity under which a detector can operate (the saturation intensity). In metrology experiments where probes can be prepared more 'cheaply' than they can be measured, it can be advantageous to distil the quantum information from many particles into a few, prior to measurement [48–50, 61, 71–74]. Ideally, all the information from a high-intensity beam of probes would be distilled into a weak beam, without the loss of any information. This feat was recently shown to be possible, with KD negativity being the enabler [49, 61].

Consider altering the general protocol outlined in figure 1. Between the unitary and the measurement, we insert a post-selective filter. The filter discards or retains an experimental trial, depending on a positive-operator-valued measure's outcome (figure 2). The measure has the form

$\mathcal{F} = \{\hat{F}_1 \equiv \hat{F}, \hat{F}_2 \equiv \hat{\mathbb{1}} - \hat{F}\}$. Only if the filter yields the $\hat{F}$ outcome is the probe sent to the final detector. Otherwise, the probe is discarded. After the post-selection, the (renormalised) quantum state is

$$|\Psi_\theta^{\mathrm{PS}}\rangle = \frac{1}{\sqrt{p_\theta^{\mathrm{PS}}}} \hat{K} |\Psi_\theta\rangle. \tag{35}$$

$p_\theta^{\mathrm{PS}} = \langle \Psi_\theta | \hat{F} | \Psi_\theta \rangle$ denotes the post-selection probability (the probe's probability of passing the filter). We have introduced a Kraus operator $\hat{K}$ such that $\hat{F} = \hat{K}^\dagger \hat{K}$.

Consider any parameterised quantum state $\hat{\rho}_\theta$. The maximum amount of Fisher information one can extract from it (by making the optimal final measurement) is the *quantum* Fisher information:

$$\mathcal{I}_Q(\hat{\rho}_\theta) = \mathrm{Tr}\left(\hat{\Gamma}_{\hat{\rho}_\theta}^2 \hat{\rho}_\theta\right) = \max_{\mathrm{meas.}}\{I(\theta|\hat{\rho}_\theta)\}. \tag{36}$$

$\hat{\Gamma}_{\hat{\rho}_\theta}$ is the symmetric logarithmic derivative, implicitly defined through $\partial_\theta \hat{\rho}_\theta = (\hat{\Gamma}_{\hat{\rho}_\theta} \hat{\rho}_\theta + \hat{\rho}_\theta \hat{\Gamma}_{\hat{\rho}_\theta})/2$ [52, 53]. The quantum Fisher information for $|\Psi_\theta^{\mathrm{PS}}\rangle$ is

$$\mathcal{I}_Q\left(\Psi_\theta^{\mathrm{PS}}\right) = 4\left(\frac{1}{p_\theta^{\mathrm{PS}}}\langle \Psi_\theta | \hat{A}\hat{F}\hat{A} | \Psi_\theta \rangle - \frac{1}{\left(p_\theta^{\mathrm{PS}}\right)^2}\left|\langle \Psi_\theta | \hat{A}\hat{F} | \Psi_\theta \rangle\right|^2\right). \tag{37}$$

The post-selected quantum Fisher information (equation (37)) can be recast in terms of an extended KD distribution. To this end, we define the 2-extended KD distribution

$$Q_{i,j,k}(\Psi_\theta) \equiv \langle \Psi_\theta | a_i \rangle \langle a_i | \hat{F}_k | a_j \rangle \langle a_j | \Psi_\theta \rangle. \tag{38}$$

The i and j indices label two instances of the eigenbasis of the generator $\hat{A}$. The k index labels the elements of the post-selective positive-operator-valued measure. To fully incorporate the post-selection into the KD distribution, we implement a quasi-probabilistic Bayesian update (see section 2.4):

$$\tilde{Q}_{i,j}^{\mathrm{PS}}(\Psi_\theta) = \frac{Q_{i,j,k=1}(\Psi_\theta)}{\sum_{i,j} Q_{i,j,k=1}(\Psi_\theta)} = \langle \Psi_\theta | a_i \rangle \langle a_i | \hat{F} | a_j \rangle \langle a_j | \Psi_\theta \rangle / p_\theta^{\mathrm{PS}}. \tag{39}$$

Equipped with the extended KD distribution $\tilde{Q}_{i,j}^{\mathrm{PS}}(\Psi_\theta)$, we can rewrite equation (37):

$$\mathcal{I}_Q\left(\Psi_\theta^{\mathrm{PS}}\right) = 4\left[\sum_{i,j} a_i a_j \tilde{Q}_{i,j}^{\mathrm{PS}}(\Psi_\theta) - \left|\sum_{i,j} a_i \tilde{Q}_{i,j}^{\mathrm{PS}}(\Psi_\theta)\right|^2\right]. \tag{40}$$

$\mathcal{I}_Q\left(\Psi_\theta^{\mathrm{PS}}\right)$ is (four times) an element of a quasi-probabilistic covariance matrix [49]. If $\tilde{Q}^{\mathrm{PS}}(\Psi_\theta)$ is a classical probability distribution, then $\mathcal{I}_Q\left(\Psi_\theta^{\mathrm{PS}}\right)$ is upper-bounded by four times the maximum variance, in any state, of $\hat{A}$:

$$\max_{\tilde{Q}_{i,j} \in [0,1]} \mathcal{I}_Q\left(\Psi_\theta^{\mathrm{PS}}\right) \leqslant 4\max_{\hat{\rho}}\left\{\mathrm{Var}_{\hat{\rho}}\left(\hat{A}\right)\right\} = (\Delta a)^2. \tag{41}$$

Δa denotes the spectral gap of $\hat{A}$.

On the other hand, if $\tilde{Q}^{\mathrm{PS}}(\Psi_\theta)$ contains negative elements, $\mathcal{I}_Q\left(\Psi_\theta^{\mathrm{PS}}\right)$ can be arbitrarily large. There is no fundamental bound on how much Fisher information can be distilled from many quantum states into a few, if the probability of obtaining information-dense states is correspondingly small. Moreover, [49, 61] show that this distillation can be approximately lossless, such that $\mathcal{I}_Q\left(\Psi_\theta^{\mathrm{PS}}\right) \times p_\theta^{\mathrm{PS}} \approx \mathcal{I}_Q(\Psi_\theta)$. In summary, the quantum Fisher information about θ, encoded in $|\Psi_\theta\rangle^{\otimes n}$, can be losslessly compressed into $|\Psi_\theta^{\mathrm{PS}}\rangle^{\otimes m}$, where m/n can be made arbitrarily small. In practice, systematic errors in the post-selective filter prevent unbounded information distillation [61]. These results can be extended to multiparameter metrology, where $\theta \to (\theta_1, \theta_2, \ldots, \theta_M)$, and to mixed states [49, 61].

The exact relation between the KD non-positivity (equation (17)) and information distillation has not been mapped out. Nevertheless, [75] derives the relation in the context of optimal post-selected metrology, which we now outline. Consider preparing a quantum system in the optimal state for probing a unitary $\hat{U}(\theta)$ parameterised by an unknown θ. Assume that the post-selection filter is optimal. The rate of Fisher-information distillation is directly proportional to the KD non-positivity (equation (17)):
$\mathcal{I}_Q\left(\Psi_\theta^{\mathrm{PS}}\right) \propto \mathcal{N}(\tilde{Q}^{\mathrm{PS}}(\Psi_\theta))$.
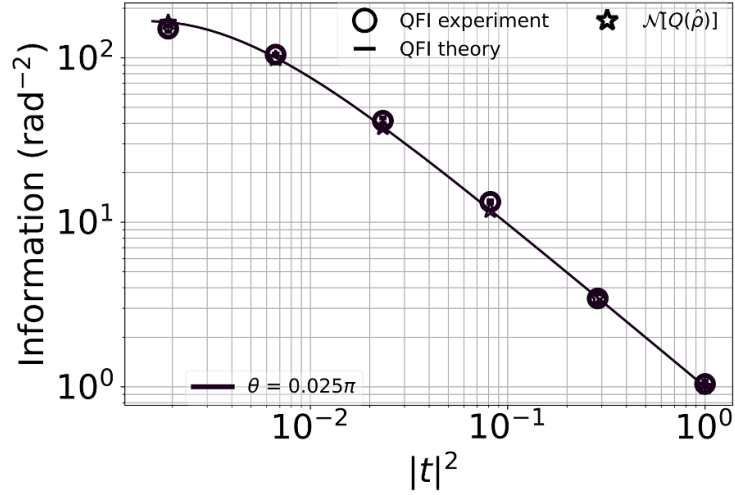
Figure 3. Post-selected quantum metrology in optics experiment. The figure is reproduced from data collected during an optical realisation [50] of figure 2. On the horizontal axis is $|t|^2$—the filter's post-selection probability, modulo small correction terms. The vertical axis shows the Fisher information, per measured photon, about an unknown birefringent phase θ. If no filtering is applied (if $|t|^2 = 1$), the maximum Fisher information (for any input state and final measurement) is $1\,\text{rad}^{-2}$ per photon. As the filtering strengthens (as $|t|^2 \to 0$), the retained photons' information content increases. The solid line shows theoretical predictions, the circles show experimental data points, and the stars show the underlying KD distribution's non-positivity (equation (17)). The stars and circles are hard to distinguish because of their overlaps. The true value of θ is 0.025π rad.

Post-selected quantum metrology has been realised experimentally. Examples of such realisations include weak-value-amplification experiments [76, 77], which we describe in section 4.3, and experiments with partially post-selected filters. In a recent proof-of-principle demonstration [50], Lupu–Gladstein *et al* used a partially post-selective filter to improve the single-photon measurement of a wave plate's birefringent phase. They increased the Fisher information per measured photon by over two orders of magnitude. Thus, the single-photon detectors could measure a low-intensity beam of single photons (a beam below the detector's saturation threshold), whilst garnering Fisher information at a high rate. Figure 3 shows data from their experiment.

4. Weak values

Several quantum-mechanical concepts involve post-selection. One of the earliest, and the most famous, example concerns *weak values* [71, 78–80]. The weak value has the form of an average of an observable, conditioned on a pre-selected initial state and a post-selected final state. Weak values are related to several concepts in the foundations of quantum mechanics (see, e.g. section 8.3.2). Below, we define the weak value and describe a protocol for measuring it (section 4.1). Then, we outline its close relation to the KD distribution (section 4.2). Finally, we describe how weak-value-measurement experiments can boost the SNR in metrological estimation of small unknown parameters (section 4.3).

4.1. Definition

We first review the theory of weak values [71, 78–80], from the perspective of von Neumann's measurement model [81]. Consider a meter that interacts with a system of interest. For simplicity, we assume that the meter interacts impulsively with the system via a unitary generated by the Hamiltonian $H_{\text{int}} = g\delta(t)\hat{p} \otimes \hat{A}$. $\hat{p}$ denotes the meter's momentum, and $\hat{A} = \sum_i a_i |a_i\rangle\langle a_i|$ is an arbitrary Hermitian system operator. The parameter g denotes the interaction strength, which controls the measurement strength. The system–meter interaction results in an entangling unitary,

$$\hat{U}_{\text{SM}} = e^{-ig\hat{p}\hat{A}}. \tag{42}$$

We define the initial system and meter state to be $|\psi_{\text{i}}\rangle|\phi\rangle$. We take the weak measurement limit, assuming that g is much less than the meter state's width (in position space). We will express the meter state in terms of the position basis $\{|x_{\text{m}}\rangle\}$. After the interaction, the joint system-meter state assumes the form

$$\int dx_{\text{m}}|x_{\text{m}}\rangle\langle x_{\text{m}}|\hat{U}_{\text{SM}}|\psi_{\text{i}},\phi\rangle = \sum_i c_i \int dx_{\text{m}}\phi(x_{\text{m}} - ga_i)|a_i\rangle|x_{\text{m}}\rangle, \tag{43}$$

where $\phi(x_m - ga_i) = \langle x_m - ga_i|\phi\rangle$ and $c_i = \langle a_i|\psi_\mathrm{i}\rangle$. The joint state is a superposition of system-and-meter products. The momentum $\hat{p}$ in equation (42) spatially translates the meter state by an amount dependent on the system state. The meter's position-basis wavefunction, $\phi(x_m - ga_i)$, encodes a shift by an amount ga_i dependent on the system state. ϕ has a width much larger than the shifts ga_i in this weak-measurement regime. The opposite regime involves a narrow ϕ: The meter state shifts by an amount large compared to the width. In this regime, textbook projective measurement occurs: By measuring the meter, one can discriminate system states unambiguously.

After the weak interaction, the system is post-selected. The weak value is defined as the average shift undergone by the meter during the post-selection. Assume that the meter's initial wavefunction, relative to the position basis, is Gaussian: $\phi(x) \propto \mathrm{e}^{-\frac{x_\mathrm{m}^2}{2\sigma^2}}$. [Although we illustrate with a Gaussian example, the following derivation holds for many forms of $\phi(x)$.] We model the system's post-selection as an arbitrary projective measurement, $\hat{\Pi}_{\psi_\mathrm{f}} = |\psi_\mathrm{f}\rangle\langle\psi_\mathrm{f}|$. The conditional meter state ϕ^PS has the form

$$
\begin{aligned}
\phi^\mathrm{PS}(x_\mathrm{m}|\psi_\mathrm{f}) &= \langle\psi_\mathrm{f}|\langle x_\mathrm{m}|\mathrm{e}^{-i g\hat{p}\hat{A}}|\psi_\mathrm{i}\rangle|\phi\rangle \\
&\approx \langle\psi_\mathrm{f}|\langle x_\mathrm{m}|\left(\hat{\mathbb{1}} - ig\hat{p}\hat{A}\right)|\psi_\mathrm{i}\rangle|\phi\rangle \\
&= \langle\psi_\mathrm{f}|\psi_\mathrm{i}\rangle\phi(x_\mathrm{m}) - ig\langle\psi_\mathrm{f}|\hat{A}|\psi_\mathrm{i}\rangle\langle x_\mathrm{m}|\hat{p}|\phi\rangle \\
&\approx \langle\psi_\mathrm{f}|\psi_\mathrm{i}\rangle\langle x_\mathrm{m}|\mathrm{e}^{-ig\hat{p}A_\mathrm{w}}|\phi\rangle \\
&= \langle\psi_\mathrm{f}|\psi_\mathrm{i}\rangle\phi(x_\mathrm{m} - gA_\mathrm{w}).
\end{aligned}
\tag{44,45}
$$

In line (44), we have Taylor-expanded to first order in g. In line (45), we have factored out the quantity $\langle\psi_\mathrm{f}|\psi_\mathrm{i}\rangle$ and re-approximated the linear-order Taylor expansion as an exponential function. The weak value of the operator $\hat{A}$ is defined as

$$
A_\mathrm{w} = \frac{\langle\psi_\mathrm{f}|\hat{A}|\psi_\mathrm{i}\rangle}{\langle\psi_\mathrm{f}|\psi_\mathrm{i}\rangle}.
\tag{46}
$$

Under the approximations above, the meter's position-basis wavefunction remains Gaussian after the post-selected weak measurement. Thus, the meter's state undergoes an update

$$
\phi(x_\mathrm{m}) \propto \mathrm{e}^{-\frac{x_\mathrm{m}^2}{2\sigma^2}} \rightarrow \phi^\mathrm{PS}(x_\mathrm{m}|\psi_\mathrm{f}) \propto \mathrm{e}^{\frac{-(x_\mathrm{m} - A_\mathrm{w})^2}{2\sigma^2}} \propto \mathrm{e}^{-\frac{[x_\mathrm{m} - g\mathrm{Re}(A_\mathrm{w})]^2}{2\sigma^2}}\mathrm{e}^{i\frac{x_\mathrm{m} g\mathrm{Im}(A_\mathrm{w})}{\sigma^2}} + \mathcal{O}(g^2).
\tag{47}
$$

One recognises, in the right-hand side, a position shift of $g\mathrm{Re}(A_\mathrm{w})$ and momentum shift $g\mathrm{Im}(A_\mathrm{w})/\sigma^2$ applied to the original wavefunction. Figure 4 provides a schematic overview of a weak-value experiment.

The weak value (equation (46)) is defined in terms of a 'pre-selected' initial state $|\psi_\mathrm{i}\rangle$ and a 'post-selected' final state $\langle\psi_\mathrm{f}|$. To facilitate the interpretation of the weak value, we let the post-selected-on state $|\psi_\mathrm{f}\rangle = |b_{j^\star}\rangle$ be one state in the basis $\{|b_j\rangle\}$ of an observable $\hat{B}$. The weak value can be rewritten as

$$
A_\mathrm{w}(\psi_\mathrm{i}, b_{j^\star}) = \frac{\langle b_{j^\star}|\hat{A}|\psi_\mathrm{i}\rangle}{\langle b_{j^\star}|\psi_\mathrm{i}\rangle} = \frac{\langle\psi_\mathrm{i}|b_{j^\star}\rangle\langle b_{j^\star}|\hat{A}|\psi_\mathrm{i}\rangle}{|\langle b_{j^\star}|\psi_\mathrm{i}\rangle|^2}.
\tag{48}
$$

To interpret the denominator, we imagine that the system is prepared in $|\psi_\mathrm{i}\rangle$, the weak interaction occurs, and $\hat{B}$ is measured. In the limit as $g \rightarrow 0$, we obtain outcome $b_{j^\star}$ with a probability $P(b_{j^\star}|\psi_\mathrm{i}) = |\langle b_{j^\star}|\psi_\mathrm{i}\rangle|^2$. The weak value satisfies the summation condition

$$
\sum_j P(b_{j^\star}|\psi_\mathrm{i}) A_\mathrm{w}(\psi_\mathrm{i}, b_{j^\star}) = \sum_j \langle\psi_\mathrm{i}|b_{j^\star}\rangle\langle b_{j^\star}|\hat{A}|\psi_\mathrm{i}\rangle = \langle\psi_\mathrm{i}|\hat{A}|\psi_\mathrm{i}\rangle.
\tag{49}
$$

Thus, the weak value $A_\mathrm{w}(b_{j^\star}, \psi_\mathrm{i})$ is a *conditioned expectation value* of $\hat{A}$.

A_w has several other interesting properties:

- A_w (equation (46)) is symmetric with respect to the exchange of $|\psi_\mathrm{i}\rangle$ and $|\psi_\mathrm{f}\rangle$, up to a complex conjugation related to the weak value's time-reversal symmetry.
- A_w can be a non-real number. As is apparent from equations (45) and (47), the real part of A_w shifts the meter's position-basis wavefunction. The imaginary part of A_w shifts the meter's conjugate, momentum-basis wavefunction [64]. Further properties of the imaginary part of A_w are discussed in [65, 66].
- $\mathrm{Re}(A_\mathrm{w})$ can lie outside the spectrum of $\hat{A}$.
- The amplitude of $\phi^\mathrm{PS}(x_\mathrm{m}|\psi_\mathrm{f})$ is reduced by the overlap between the system's initial and final states [82]. Post-selecting on a final state re-normalises the amplitude.

4.2. Connection between weak values and the KD quasi-probability distribution

Whilst the above-outlined explanation of the weak value is commonly found in the literature, the connection to the KD distribution is not obvious. To further unpack the weak value's structure and reveal the connection to the KD distribution, we recall the spectral expansion $\hat{A} = \sum_i a_i |a_i\rangle\langle a_i|$. Substituting it into equation (48), we expand the weak value to

$$A_{\mathrm{w}}(\psi_{\mathrm{i}}, b_{j^\star}) = \frac{\langle b_{j^\star}|\hat{A}|\psi_{\mathrm{i}}\rangle}{\langle b_{j^\star}|\psi_{\mathrm{i}}\rangle} = \sum_i a_i \frac{\langle b_{j^\star}|a_i\rangle\langle a_i|\psi_{\mathrm{i}}\rangle}{\langle b_{j^\star}|\psi_{\mathrm{i}}\rangle} = \sum_i a_i \frac{\langle b_{j^\star}|a_i\rangle\langle a_i|\psi_{\mathrm{i}}\rangle\langle\psi_{\mathrm{i}}|b_{j^\star}\rangle}{|\langle b_{j^\star}|\psi_{\mathrm{i}}\rangle|^2}$$

$$= \sum_i a_i \frac{Q_{i,j^\star}(\psi_{\mathrm{i}})}{P(b_{j^\star}|\psi_{\mathrm{i}})} = \sum_i a_i \tilde{Q}_{i|j^\star}(\psi_{\mathrm{i}}). \tag{50}$$

We have defined the KD distribution $Q_{i,j}(\psi_{\mathrm{i}}) \equiv \langle b_j|a_i\rangle\langle a_i|\psi_{\mathrm{i}}\rangle\langle\psi_{\mathrm{i}}|b_j\rangle$. Conditioning on $j = j^\star$ yields $\tilde{Q}_{i|j^\star}(\psi_{\mathrm{i}})$, in accordance with equation (25). The last equality in equation (50) shows that the weak value is an average of the $\hat{A}$ eigenvalues with respect to a conditional quasi-probability distribution [83]. Marginalising the joint distribution $Q_{i,j}(\psi_{\mathrm{i}})$ over i yields the post-selection probability. Furthermore, suppose that $\hat{A}$ is a rank-1 projector: $\hat{A} = |a_1\rangle\langle a_1|$. The weak value equals a conditional KD quasi-probability: $\hat{A}_{\mathrm{w}}(\psi_{\mathrm{i}}, b_{j^\star}) = \tilde{Q}_{i=1|j^\star}(\psi_{\mathrm{i}})$.

An intuitive *operational meaning* of the weak value is a conditioned average of the outcomes of weak measurements of the observable $\hat{A}$. This intuition is supported by the weak value's formal equivalence to both a conditioned expectation value and a conditioned quasi-probabilistic average of the $\hat{A}$ eigenvalues. Conditioning a von Neumann measurement in the weak-coupling (small-disturbance) limit yields the KD distribution as the joint distribution over the outcomes of the two ordered measurements performed. Moreover, the last equality in equation (50) highlights how $A_{\mathrm{w}}(\psi_{\mathrm{i}}, b_{j^\star})$ lies between the minimum and maximum $\hat{A}$ eigenvalue if $Q_{i,j}(\psi_{\mathrm{i}})$ is a classical probability distribution. However, if $Q_{i,j}(\psi_{\mathrm{i}})$ is non-positive, $A_{\mathrm{w}}(\psi_{\mathrm{i}}, b_{j^\star})$ can lie outside this range. In some experiments, such *anomalous* weak values witness the non-classical phenomenon of contextuality (section 8.3.2). Moreover, anomalous weak values have been used to amplify SNRs in experiments (see section 4.3).

Importantly, the connection between conditional KD quasi-probabilities and measurable weak values allows the quasi-probabilities to be experimentally measured. Thus, despite their anomalous behaviour, KD quasi-probabilities are empirically meaningful as testable predictions. Moreover, by equation (25), a quantum state's KD representation decomposes into conditional quasi-probabilities

$$Q_{i,j}(\psi_{\mathrm{i}}) = \tilde{Q}_{i|j}(\psi_{\mathrm{i}}) P(b_j|\psi_{\mathrm{i}}). \tag{51}$$

Consequently, to infer $Q_{i,j}(\psi_{\mathrm{i}})$ one can measure $\tilde{Q}_{i|j}(\psi_{\mathrm{i}})$ and $P(b_j|\psi_{\mathrm{i}})$ independently. As we saw above, we can infer $\tilde{Q}_{i|j}(\psi_{\mathrm{i}})$ and $P(b_j|\psi_{\mathrm{i}})$ via a weak measurement of $\hat{A} = |a_i\rangle\langle a_i|$, by initialising the system in $|\psi_{\mathrm{i}}\rangle$ and post-selecting on $|b_j\rangle$. To infer an entire KD distribution $Q(\psi_{\mathrm{i}})$, one can perform a set of weak-measurement experiments, scanning through the values of i and j. If $\langle a_i|b_j\rangle \neq 0$ for all i and j, then $\tilde{Q}(\psi_{\mathrm{i}})$ is informationally complete about $|\psi_{\mathrm{i}}\rangle$ (equation (15)). This quasi-probability method for measuring quantum-state amplitudes has allowed states to be determined and even experimentally tracked through successive dynamical updates via Bayes' rule [47]. Thus, the connection to measurable weak values elevates the KD distribution from an abstract state representation to an experimentally significant and practical representation. Section 5 contains further details about the KD distribution in relation to quantum-state measurements.

4.3. Weak-value amplification

Perhaps the best-known application of weak values is weak-value amplification (figure 4). Consider the weak-value experiment discussed in section 4.1. Assume that the interaction strength g is small and unknown. Weak-value amplification improves measurements of g. The fact that the meter degree of freedom can be shifted arbitrarily far may be viewed as an amplified response of the combined system. If the resulting signal is re-scaled by the weak value, the parameter g can be measured precisely. This technique is now widely used and is applied to many record-breaking metrology experiments [76, 77, 84].

Weak-value amplification falls under the above-reviewed topic of post-selected metrology (section 3.3). However, in its simplest form, weak-value amplification is most easily understood not in terms of Fisher information, but in terms of the SNR.

We can compare the final meter state (described in section 4.1) realised in experiments that involve post-selection and experiments that do not. We can estimate g by measuring the meter's position on a detector. We assume that experimental repetitions are uncorrelated. The SNR quantifies the estimation's sensitivity:

Figure 4. Weak-value amplification. In (a) and (b), the system particle is initialised in an eigenstate corresponding to the minimum or maximum eigenvalue of $\hat{A}$. Initially, the meter particle is moving. The meter has a spatial degree of freedom associated with an axis perpendicular to the motion. That degree of freedom is prepared in a state represented, relative to the position eigenbasis, by a Gaussian. The system interacts with the meter. A screen detector then measures the meter's position wavefunction. The difference between the maxima in (a) and (b) can be used to estimate the coupling strength *g*. In (c), the system is prepared in $|\psi_i\rangle$, and $\langle\psi_f|$ is post-selected on. Only if the post-selection succeeds is the meter measured. Then, the shift in the meter's position can lie outside the range defined by the spectral gap of $\hat{A}$. We have re-normalised the position profile in (c) by the post-selection probability.

$$\mathcal{R}^{\mathrm{SNR}} = \sqrt{N}\frac{|\langle x_{\mathrm{m}}\rangle|}{\sqrt{\mathrm{Var}\,(x_{\mathrm{m}})}}. \tag{52}$$

$\mathrm{Var}(x_{\mathrm{m}}) = \langle x_{\mathrm{m}}^2\rangle - \langle x_{\mathrm{m}}\rangle^2$ is the variance in the position-measurement outcomes. $N \gg 1$ denotes the number of experimental trials. $\mathcal{R}^{\mathrm{SNR}}$ quantifies our ability to distinguish the detected signal's mean from the measurement noise.

We can apply these concepts to the weak-value-amplification protocol and contrast it with a weak-value-free alternative. Let us return to the state in equation (43). Let the system be prepared in the eigenstate associated with the maximum *A* eigenvalue $a_{\mathrm{max}} = 1$. Let the meter distribution, be Gaussian with a width $\langle x_{\mathrm{m}}^2\rangle = \sigma^2$. The SNR has the form

$$\mathcal{R}_{\mathrm{s}}^{\mathrm{SNR}} = \frac{\sqrt{N}g}{\sigma}. \tag{53}$$

We have defined *N* as the number of weak-measurement trials (figure 4). The trials begin with the same initial system and meter states. We assume that all meter–system interactions have the same strength *g*, for simplicity.

In contrast, consider the weak-value-amplification protocol. For simplicity, suppose that the weak value is real and positive: $A_{\mathrm{w}} = \mathrm{Re}(A_{\mathrm{w}}) > 0$. The SNR almost has the form of (53). However, we must replace *g* with $A_{\mathrm{w}}g$ and *N* with N_{PS}, the number of trials that pass the post-selection (the meter Gaussian's width remains unchanged in the simplest weak-value amplification):

$$\mathcal{R}_{\mathrm{wv}}^{\mathrm{SNR}} = \frac{\sqrt{N_{\mathrm{PS}}}A_{\mathrm{w}}g}{\sigma}. \tag{54}$$

By equation (47), the meter wavefunction's peak is shifted by an amount $A_{\mathrm{w}}g \gg g$, exhibiting signal amplification. The meter's final wave function has the same width as its initial wave function. The post-selection's probabilistic nature implies that the number N_{PS} of data points, collected from *N* trials of the experiment, will vary from batch of trials to batch of trials, if we run multiple batches of *N* trials. We calculate the post-selected SNR by replacing N_{PS} with the expected number of data points, $N|\langle\psi_f|\psi_i\rangle|^2 + \mathcal{O}(g^2)$, the post-selection probability times the number of trials (to within a small correction). The SNR becomes

$$\mathcal{R}_{\mathrm{wv}}^{\mathrm{SNR}} = \mathcal{R}_{\mathrm{s}}^{\mathrm{SNR}}\langle\psi_f|\hat{A}|\psi_i\rangle = \mathcal{R}_{\mathrm{s}}^{\mathrm{SNR}}\langle\psi_f|\hat{A}|\psi_i\rangle. \tag{55}$$

In many cases, the coefficient $\langle\psi_f|\hat{A}|\psi_i\rangle$ of proportionality can be made to lie close to 1, matching the standard measurement's SNR. This calculation was first published (to the best of our knowledge) in [85].

The conclusion of this calculation is that, at best, weak-value amplification achieves the standard technique's precision. However, something remarkable has happened: we can obtain the same precision as in the standard case, but by performing final measurements in a tiny fraction $N_{\mathrm{PS}} \ll N$ of the trials. This calculation relied on an assumption: Initiating trials is easy, whereas the final measurement is a valuable resource. Often in optics, for example, the limiting resource is not the amount of power that a laser can emit, but the amount of power that a detector can receive. From this perspective, if we regard the number of detected photons as the limiting resource, weak-value amplification achieves a huge advantage: Consider matching the number of photons detected in the standard scheme to the number detected in the weak-value-amplification scheme. We boost the SNR by a factor equal to the weak value, $A_{\mathrm{w}} \gg 1$.

Additionally, weak-value amplification offers many other advantages. Some are:

- Often, the weak-value experiment's setup is more robust with respect to technical noise than a standard experiment's setup is. A simple example involves optical-beam-deflection measurements. In such measurements, slight turbulence can limit the precision of the measurement of a beam-deflection angle. The standard measurement technique focuses down an optical beam, to reduce the surface area of the exposed transverse region of the detector. A slight beam deflection translates the beam laterally by the greatest possible amount, relative to the beam profile, giving the measurement the greatest sensitivity. The weak-value method (in this case, relying on the weak value's imaginary part) makes the beam profile as large as possible. In principle, both techniques offer the same shot-noise-limited precision. However, the weak-value-amplification method vastly outperforms the standard method in the presence of turbulence [86].

- If the system suffers from *systematic error*, averaging over more trials does not suppress it. The reason is that systematic error affects accuracy, not precision. Weak-value amplification can benefit also this situation [86–90]. Consider amplifying the signal by the weak value. Dividing by the weak value, to estimate the parameter g, can suppress the systematic error by the weak value. A simple example concerns bias offset. Suppose that we aim to measure the origin g of a meter. Let the meter's true origin differ from the assumed origin by a small amount ε. Consider averaging over measurement outcomes from the set $\{x_m^{(i)}\}$. No matter how much data you have, the average will deviate from the true g value by ε, which affects each measurement. Under the weak-value-amplification scheme, however, the measured average is $A_w g$, which has the same error, ε. Consider dividing the measured data by A_w to estimate g. The systematic error is reduced from ε to $\varepsilon/A_w \ll \varepsilon$, improving the accuracy. We stress that this reduction is independent of the amount of data. However, if a systematic error obscures the to-be-measured parameter g, the weak-value-amplification method will not help: the systematic error will be amplified along with g.

- In weak-value amplification, one post-selects the data that satisfy the post-selection criterion. Yet you need not discard the remaining particles, which need not even be measured destructively. They can be reused in another context [91] or recycled. Indeed, optimal experiments have realised the following outcome: Even in the ideal case, free of technical noise, re-injecting non-post-selected photons into an optical interferometer has improved the SNR [92–95].

5. The KD distribution and direct measurements of quantum states

We now discuss ways of measuring a quantum state's KD distribution, focusing on *direct measurement*. Means of experimentally determining quantum states has practical and foundational applications. The problem's history encompasses contributions by Birkhoff and von Neumann [96], Fano [97], Pauli [98] and others. *Prima facie*, the quantum state seems abstract—a complex-valued amplitude distribution, of which one could glimpse only limited features in a measurement. Not until the 1990s was quantum tomography invented [99] and demonstrated [100]. The first studies centred on a light wave's quantum state, but tomography of other systems followed quickly.

    The term *quantum tomography* is sometimes used to mean quantum-state determination. However, it is actually a specific procedure. It involves measuring in diverse bases (of the Hilbert space directly associated with a system) that, together, span the density-operator space. Performing these measurements on an ensemble of identical systems, one can infer real-valued probability distributions that resemble shadows of the complex state. From these distributions, one can reconstruct the state [101][11]. A measurement that directly outputs the KD distribution (up to a normalisation factor) removes the need for such a reconstruction. We show that this elimination has practical and fundamental implications.

    Direct measurements of quantum states are used to infer several quasi-probability distributions, not only the KD distribution. Direct measurement comprises a wide range of measurement procedures, which can be 'direct' in some or all of these senses:

1. The procedure does not require a complicated mathematical reconstruction (as described just above).
2. The procedure is local: it measures the state's quasi-probability amplitude [e.g. $Q_{i,j}(\hat{\rho})$] at a given point in phase space (a_i, b_j).
3. That amplitude's value appears directly on a measurement apparatus.
4. The experimental procedure is simple and general. For example, one might measure $\hat{A}$ and then $\hat{B}$. Tomography, in contrast, requires many measurements, potentially in exotic bases that are not easily accessible experimentally.

---

[11] The word *tomography* stems from *tomos* (slice) and *graphy* (writing). In medicine, two-dimensional x-ray shadows (slices), taken at different angles, are used in the computer-aided tomographic reconstruction (a CAT scan) of a three-dimensional object—for example, a skull.

Direct measurement based on weak measurement has all these properties. The laws of quantum physics (e.g. the no-cloning theorem [102]) require that, to determine an arbitrary quantum state, we must measure an ensemble of identical systems. Thus all direct-measurement procedures involve averages over measurement outcomes from an ensemble, similarly to quantum tomography. However, direct measurement offer advantages over tomography.

Direct measurements of discrete and continuous KD distributions determine quantum states. The most prominent results have concerned direct measurements of continuous-variable quantum states. We summarised a few properties of the continuous KD distribution in section 2.3. Now, in section 5.1, we outline how to measure quantum states directly. We describe two strategies: One involves the Husimi distribution and no weak measurements, and the other involves weak measurements and the KD distribution. In sections 5.2 and 5.3, we review generalisations of, and other procedures, for direct measurements.

5.1. Direct quantum-state measurements

Measuring a classical point particle's position and momentum directly determines its state. In quantum physics, one cannot precisely and simultaneously measure a particle's position (equation (18)) and momentum (equation (19)). Nor can one even ascribe precise values to $\hat{x}$ and $\hat{p}$ simultaneously. One can measure position and momentum simultaneously [103], if the information obtained about both quantities is imprecise. Consider impreciseness in the form of uncertainty: $x \pm \Delta x$. [104] concerns minimal measurement-device uncertainty, which is balanced between simultaneous measurements of position and momentum: $\Delta x = \Delta p$. The outcome (x', p') corresponds to a projection onto a coherent state $|\alpha' = x' + ip'\rangle$. This outcome's probability is proportional to the Husimi quasi-probability distribution: $\mathbb{P}(x', p') = \mathrm{Tr}\left(|\alpha'\rangle\langle\alpha'|\hat{\rho}\right) \propto H_{x',p'}(\hat{\rho})$. The Husimi distribution $H_{x,p}(\hat{\rho})$ is a faithful representation of the quantum state $\hat{\rho}$ [104]. In this sense, one can extract information about the quantum system, as from a classical particle, at a phase-space point (x', p'), using imprecise measurements. One thereby directly measures the state.

Similarly to the Husimi distribution, the KD distribution corresponds to the measurement of a point in phase space. However, the continuous-variable KD distribution $Q(x, p; \psi)$ (see section 2.3) does not correspond to projections onto a state such as $|\alpha\rangle$, which has finite uncertainties, Δx and Δp. Recall the rightmost side of equation (22),

$$Q(x, p; \psi) = \mathrm{Tr}\left(\hat{\pi}_p \hat{\pi}_x \hat{\rho}\right), \tag{56}$$

wherein $\hat{\pi}_x \equiv |x\rangle\langle x|$ and $\hat{\pi}_p \equiv |p\rangle\langle p|$. This equation appears to encode an expectation value of $\hat{\pi}_p \hat{\pi}_x$. This observation suggests that the KD distribution corresponds to projecting onto a position eigenstate $|x\rangle$ and a momentum eigenstate $|p\rangle$ (such that $\Delta x = \Delta p = 0$), at least intuitively. However, $\hat{\pi}_p \hat{\pi}_x$ is not Hermitian and thus not an observable. If it were an observable, a measurement of $\hat{\pi}_p \hat{\pi}_x$ would violate uncertainty principles, as it would simultaneously fix position and momentum. But a direct measurement of the KD distribution need only output the *expectation value* of $\hat{\pi}_p \hat{\pi}_x$—output $\mathrm{Tr}\left(\hat{\pi}_p \hat{\pi}_x \hat{\rho}\right)$. This value need not be determined from a precise measurement in a particular trial. Rather, one can determine the value by averaging the imprecise results of a sequence of measurements performed on an ensemble of identically prepared quantum systems. In this sense, as we discuss next, the non-Hermitian operator $\hat{\pi}_p \hat{\pi}_x$ can be can be 'measured'. In turn, the KD distribution can be directly measured experimentally.

One way to measure the average of $\hat{\pi}_p \hat{\pi}_x$, and directly determine the KD distribution, is to use weak measurements. (For details about weak measurements, see section 4.) Weakly measuring $\hat{\pi}_x$ leaves a quantum state $\hat{\rho} = |\psi\rangle\langle\psi|$ mostly unchanged. Therefore, one can perform a subsequent strong (ordinary) measurement of $\hat{\pi}_p$, obtaining information about the initial state. Post-selected on a momentum value p (in a sub-ensemble of the ensemble of all momentum states), the position-projector weak measurement's average result is the weak value, equation (50):

$$\pi_{x\mathrm{w}}(\psi, p) = \frac{\langle p|\hat{\pi}_x|\psi\rangle}{\langle p|\psi\rangle} = \frac{\langle p|\hat{\pi}_x|\psi\rangle\langle\psi|p\rangle}{|\langle p|\psi\rangle|^2} \equiv \tilde{Q}(x; \psi|p). \tag{57}$$

(We have reproduced equation (46), with $\hat{A} = \hat{\pi}_x$ and $\langle\psi_{\mathrm{f}}| = \langle p|$.) In what follows, we denote the system's momentum and position by x and p. We denote the meter's position by x_{m}. As outlined in section 4.2, the conditional KD distribution is related to an observable's weak value. When the observable is a projector, as here, the weak value equals a pure state's conditional KD-distribution density: $\tilde{Q}(x; \psi|p)$.

There is an even tighter connection between the KD distribution and weak measurements. Consider performing a weak measurement, without post-selecting on the sub-ensemble that realises outcome p. This non-conditioned weak measurement's average result is sometimes called the *weak average*. We denote it by $\langle\hat{\pi}_p \hat{\pi}_x\rangle_{\mathrm{w}}$ and define it below, in equations (58) and (59). Recall that the weak value probes whether the

system first had the position x, *given that* it subsequently had momentum p. In contrast, the weak average probes whether the system had properties *x and* then p. In section 4.1, we derived an expression (equation (47)) for a meter's late-time position-basis wavefunction. The meter's wavefunction was initially a Gaussian, and then the system was pre- and post-selected. We can apply this result to calculate the meter's final spatial probability density in a weak-value experiment intended to measure $\pi_{xw}(\psi,p)$: $|\phi^{\mathrm{PS}}(x_{\mathrm{m}}|p)|^2 \propto \exp\left([x_{\mathrm{m}} - g\mathrm{Re}(\pi_{xw}(\psi,p))]^2/\sigma^2\right) + \mathcal{O}(g^2)$. Again, g denotes the weak measurement's strength. In other words, $\mathbb{P}(x_{\mathrm{m}}|p) \equiv |\phi^{\mathrm{PS}}(x_{\mathrm{m}}|p)|^2$ is the probability density associated with the meter's being at position x_{m}, if the system's momentum was post-selected on being p. For simplicity, we assume that the weak value is real. One experimentally infers the weak value from the shifted meter's mean final position, $\bar{x}_{\mathrm{m}}$ (see section 4.1 and figure 4). The average is calculated over the conditional joint system–meter measurement distribution $\mathbb{P}(x_{\mathrm{m}}|p)$. That is, $\frac{1}{g}\pi_{xw}(\psi,p) \approx \bar{x}_{\mathrm{m}} = \langle \hat{x}_{\mathrm{m}} \rangle = \int x_{\mathrm{m}}\,\mathbb{P}(x_{\mathrm{m}}|p)\,\mathrm{d}x_{\mathrm{m}}$. The weak average manifests similarly but depends on the unconditional joint distribution, $\mathbb{P}(x_{\mathrm{m}},p) \equiv \mathbb{P}(x_{\mathrm{m}}|p)\,\mathbb{P}(p) = |\phi^{\mathrm{PS}}(x_{\mathrm{m}}|p)|^2 |\langle p|\psi \rangle|^2$. That is, $\langle \hat{\pi}_p \hat{\pi}_x \rangle_{\mathrm{w}} \propto \langle \hat{x}_{\mathrm{m}} \hat{\pi}_p \rangle = \int x_{\mathrm{m}}\,\mathbb{P}(x_{\mathrm{m}},p)\,\mathrm{d}x_{\mathrm{m}}$. Consequently, the weak average is

$$\langle \hat{\pi}_p \hat{\pi}_x \rangle_{\mathrm{w}} \equiv \pi_{xw}(\psi,p) \cdot \mathbb{P}(p) = \langle p|\hat{\pi}_x|\psi \rangle \langle \psi|p \rangle = \mathrm{Tr}\left(\hat{\pi}_p \hat{\pi}_x |\psi \rangle \langle \psi| \right). \tag{58}$$

The generalisation beyond pure states follows from the linearity of $\hat{\rho}$:

$$\langle \hat{\pi}_p \hat{\pi}_x \rangle_{\mathrm{w}} \equiv \mathrm{Tr}\left(\hat{\pi}_p \hat{\pi}_x \hat{\rho} \right) = Q(x,p;\hat{\rho}). \tag{59}$$

These results were proved in [105, 106]. In summary, a weak measurement of $\hat{\pi}_x$, followed by strong measurement of $\hat{\pi}_p$, directly yields the KD representation of the state.
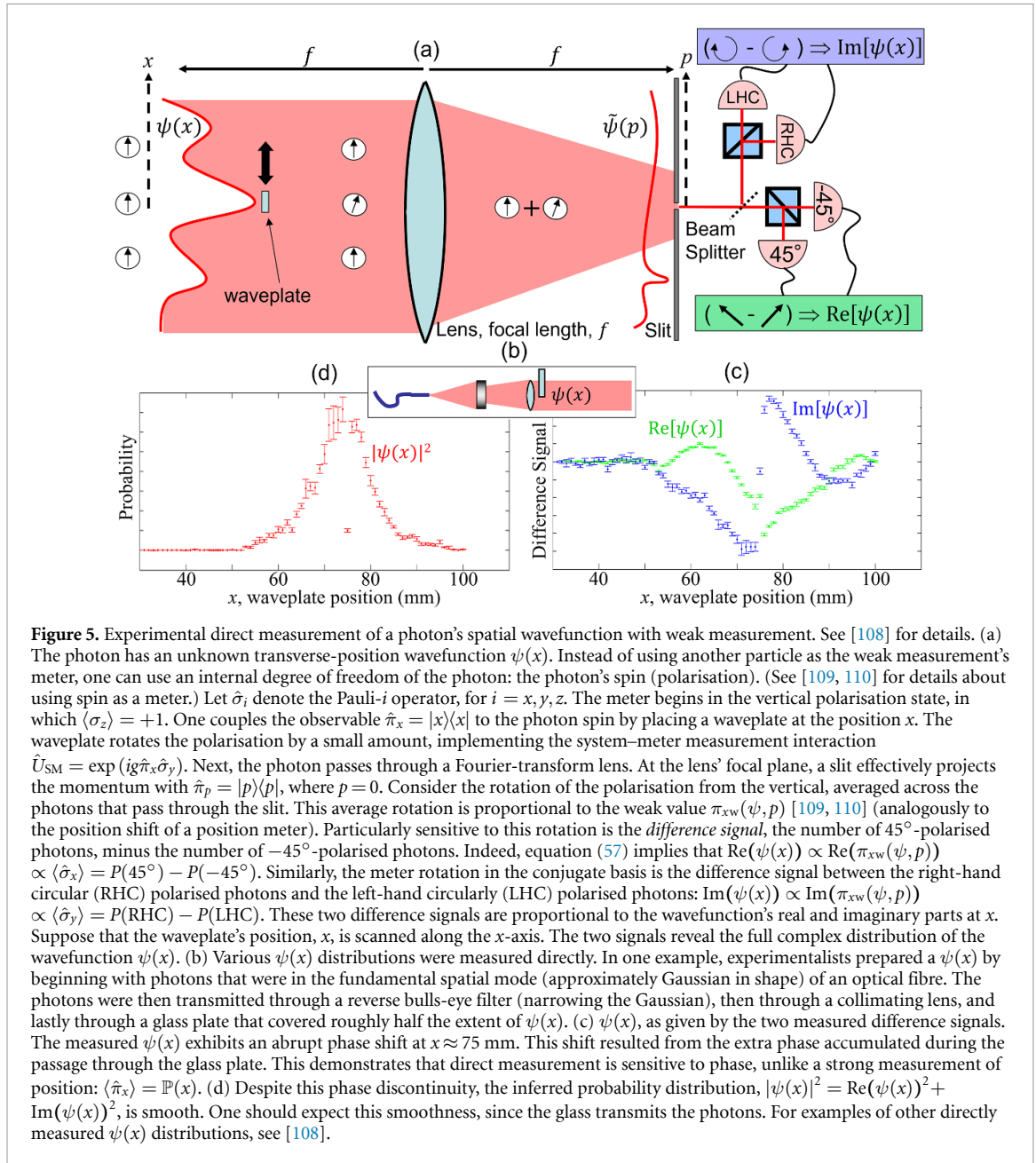
How, then, are weak measurements enabling the simultaneous measurement of position and momentum? As alluded to above, one must forfeit precision. However, unlike in the direct measurement of the Husimi distribution, the imprecision must not be in the form of position or momentum uncertainty, Δx or Δp. After all, we are now nominally measuring $\hat{\pi}_p$ and $\hat{\pi}_x$ eigenstates, which respectively have zero momentum and position uncertainties. The answer was presented in [107]. It comes from the fact that, in the weak-measurement regime, the average meter shift is small, compared to the meter's original position uncertainty (i.e. the system-meter coupling g is small). Consequently, in any given trial, little information is gained from a weak measurement: weak measurements are inherently imprecise. For example, measuring a system observable $\hat{\pi}_x$ weakly reduces a type of precision called *predictability*, $\mathbb{P}(x|x_{\mathrm{m}})$ [107]. (Recall that x denotes the system's position and x_{m} denotes the meter's position.) If x and x_{m} are such that $\mathbb{P}(x|x_{\mathrm{m}}) \ll 1$, the meter's position reveals little about the system's position in any one trial. Only by averaging over many trials can the reduction in per-trial information be overcome and can $\langle \hat{\pi}_p \hat{\pi}_x \rangle_{\mathrm{w}} = Q(x,p;\hat{\rho})$ be determined with little statistical uncertainty.

The direct measurement of the quantum wavefunction was first proposed as a process involving weak measurements. It was first experimentally demonstrated for a photon's transverse spatial wavefunction $\psi(x)$ (figure 5) [108]. If one weakly measures $\hat{\pi}_x$ and post-selects on $p=0$, equation (57) gives the weak value: $\pi_{xw}(\psi,p{=}0) = \langle p{=}0|x \rangle \langle x|\psi \rangle / \langle p{=}0|\psi \rangle = \psi(x)/[\psi_{\mathrm{f}}(0)\sqrt{2\pi\hbar}] = K\psi(x)$. We have combined the constant factors into K. Thus, the weak value is proportional to the complex wavefunction. Unlike in quantum tomography, however, the wavefunction appears directly on the measurement apparatus here. In the experiment, the photon's spin degree of freedom (polarisation) served as a meter. The spin was coupled to the photon's transverse position x as follows with help from a small waveplate. The waveplate was placed so as, at a chosen position x, to slightly rotate the photon's linear polarisation. Then, a Fourier-transform lens and slit measured $\hat{\pi}_{p=0}$. As the waveplate was moved along x, the weak value was recorded for each value of x, giving the real and imaginary parts of $\psi(x)$. In this sense, one can directly observe the complex wavefunction associated with an ensemble of identically prepared quantum particles, up to the constant normalisation factor K.

5.2. Generalisations of direct measurement

Soon after the first demonstration of direct measurement, generalisations began appearing. Different photon quantum states have been measured directly; examples include the polarisation [111–113], 2D-spatial [114], orbital-angular-momentum [115, 116] and time-frequency states [117]. Non-photonic systems have also been measured directly; examples include matter waves [118] and a nuclear-magnetic-resonance quantum processor [119].

Alternatively, one can generalise the post-selection. Consider measuring the weak average for every p value, rather than solely at $p=0$. (One could, for example, use a camera instead of a slit.) Then, as equation (59) shows, one would directly measure the KD distribution, rather than the wavefunction [106].

Figure 5. Experimental direct measurement of a photon's spatial wavefunction with weak measurement. See [108] for details. (a) The photon has an unknown transverse-position wavefunction $\psi(x)$. Instead of using another particle as the weak measurement's meter, one can use an internal degree of freedom of the photon: the photon's spin (polarisation). (See [109, 110] for details about using spin as a meter.) Let $\hat{\sigma}_i$ denote the Pauli-i operator, for $i = x, y, z$. The meter begins in the vertical polarisation state, in which $\langle \sigma_z \rangle = +1$. One couples the observable $\hat{\pi}_x = |x\rangle\langle x|$ to the photon spin by placing a waveplate at the position x. The waveplate rotates the polarisation by a small amount, implementing the system–meter measurement interaction $\hat{U}_{\text{SM}} = \exp\left(ig\hat{\pi}_x\hat{\sigma}_y\right)$. Next, the photon passes through a Fourier-transform lens. At the lens' focal plane, a slit effectively projects the momentum with $\hat{\pi}_p = |p\rangle\langle p|$, where $p = 0$. Consider the rotation of the polarisation from the vertical, averaged across the photons that pass through the slit. This average rotation is proportional to the weak value $\pi_{x\text{w}}(\psi, p)$ [109, 110] (analogously to the position shift of a position meter). Particularly sensitive to this rotation is the *difference signal*, the number of $45°$-polarised photons, minus the number of $-45°$-polarised photons. Indeed, equation (57) implies that $\text{Re}\big(\psi(x)\big) \propto \text{Re}\big(\pi_{x\text{w}}(\psi, p)\big)$ $\propto \langle\hat{\sigma}_x\rangle = P(45°) - P(-45°)$. Similarly, the meter rotation in the conjugate basis is the difference signal between the right-hand circular (RHC) polarised photons and the left-hand circularly (LHC) polarised photons: $\text{Im}\big(\psi(x)\big) \propto \text{Im}\big(\pi_{x\text{w}}(\psi, p)\big)$ $\propto \langle\hat{\sigma}_y\rangle = P(\text{RHC}) - P(\text{LHC})$. These two difference signals are proportional to the wavefunction's real and imaginary parts at *x*. Suppose that the waveplate's position, *x*, is scanned along the *x*-axis. The two signals reveal the full complex distribution of the wavefunction $\psi(x)$. (b) Various $\psi(x)$ distributions were measured directly. In one example, experimentalists prepared a $\psi(x)$ by beginning with photons that were in the fundamental spatial mode (approximately Gaussian in shape) of an optical fibre. The photons were then transmitted through a reverse bulls-eye filter (narrowing the Gaussian), then through a collimating lens, and lastly through a glass plate that covered roughly half the extent of $\psi(x)$. (c) $\psi(x)$, as given by the two measured difference signals. The measured $\psi(x)$ exhibits an abrupt phase shift at $x \approx 75$ mm. This shift resulted from the extra phase accumulated during the passage through the glass plate. This demonstrates that direct measurement is sensitive to phase, unlike a strong measurement of position: $\langle\hat{\pi}_x\rangle = \mathbb{P}(x)$. (d) Despite this phase discontinuity, the inferred probability distribution, $|\psi(x)|^2 = \text{Re}\big(\psi(x)\big)^2 + \text{Im}\big(\psi(x)\big)^2$, is smooth. One should expect this smoothness, since the glass transmits the photons. For examples of other directly measured $\psi(x)$ distributions, see [108].
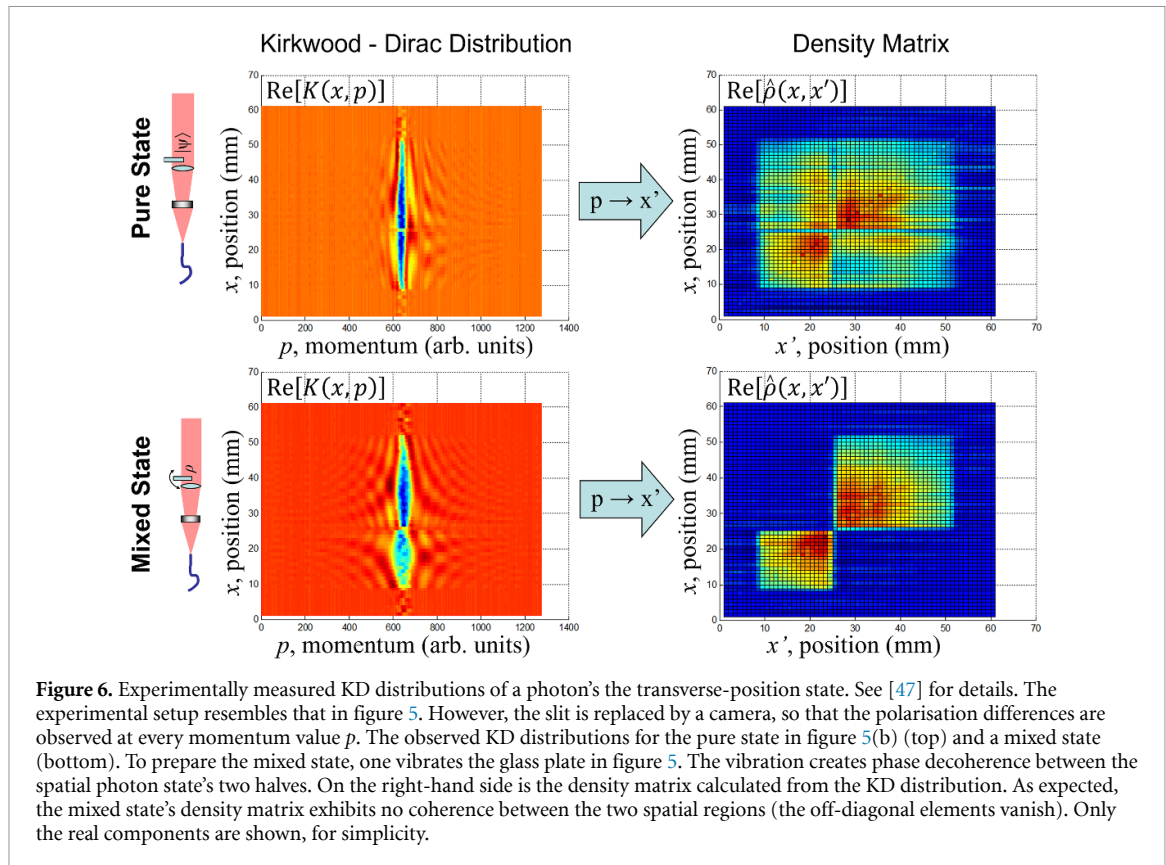
Further work has established a relationship between weak values and probability distributions over simultaneous measurements' possible outcomes [120].

Moreover, direct measurement generalises straightforwardly to discrete quantum states. Hence the KD representation of a photon's polarisation state was soon measured directly [111]. Using all *p* enables the direct measurement of mixed quantum states.

A central concept in imaging and optical coherence is a photon's transverse position. One of the first determinations, by any method, of the mixed state of a photon's transverse position was accomplished via direct measurement of the KD distribution [47]. Figure 6 shows the procedure, which leveraged weak measurement. By extending the concept to a sequence of *k* weak measurements, [106] showed that a *k*-extended KD distribution could be measured. [106] reported on a sequence of three weak measurements of projectors in two complementary bases, $\{|a_i\rangle\}$ and $\{|b_k\rangle\}$. The following density matrix was obtained:

$$\rho_{a_i, a_j} \equiv \langle a_i | \hat{\rho} | a_j \rangle \propto \left\langle \hat{\pi}_{a_i}\hat{\pi}_{b_k}\hat{\pi}_{a_j} \right\rangle_{\text{w}} \equiv \text{Tr}\left(\hat{\pi}_{a_i}\hat{\pi}_{b_k}\hat{\pi}_{a_j}\hat{\rho}\right) \equiv Q_{i,k,j}(\hat{\rho}). \tag{60}$$

References [121, 122] reported on the measurement of the real parts of weak values of two-photon entangled states. Although these papers' procedure and results corresponded to direct measurement, the paper predate the concept and, thus, do not refer to it. Not until [123] were entangled states' full complex amplitudes explicitly measured directly. In summary, diverse states and systems can be measured directly.
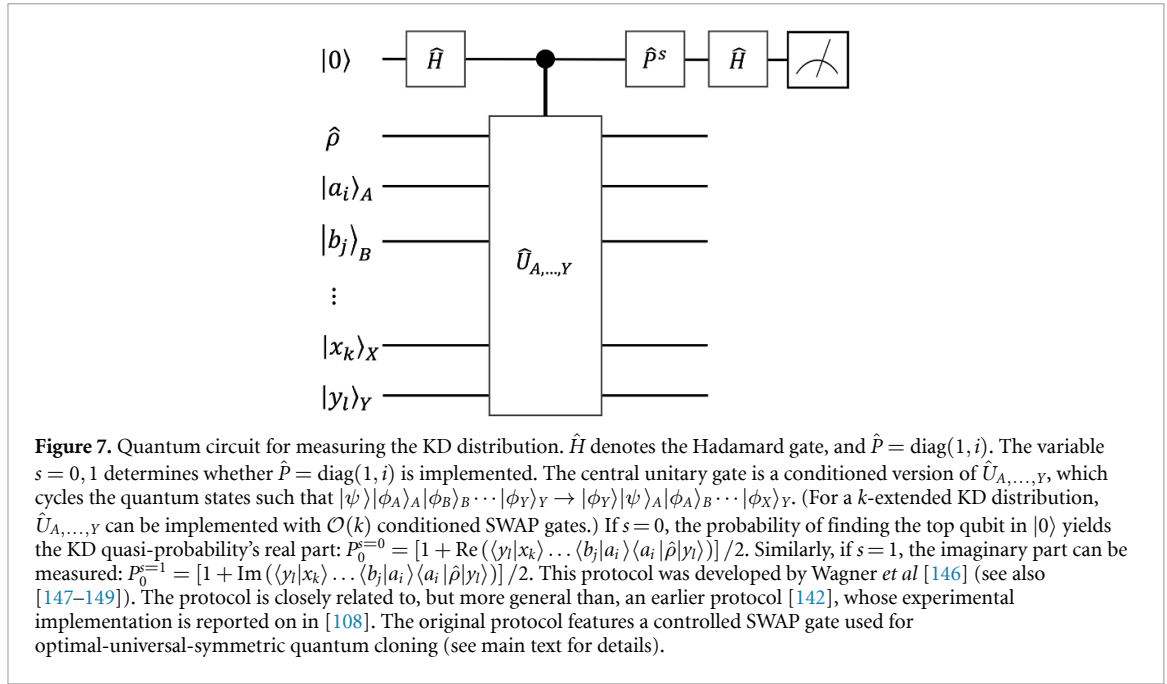
Figure 6. Experimentally measured KD distributions of a photon's the transverse-position state. See [47] for details. The experimental setup resembles that in figure 5. However, the slit is replaced by a camera, so that the polarisation differences are observed at every momentum value *p*. The observed KD distributions for the pure state in figure 5(b) (top) and a mixed state (bottom). To prepare the mixed state, one vibrates the glass plate in figure 5. The vibration creates phase decoherence between the spatial photon state's two halves. On the right-hand side is the density matrix calculated from the KD distribution. As expected, the mixed state's density matrix exhibits no coherence between the two spatial regions (the off-diagonal elements vanish). Only the real components are shown, for simplicity.

On the practical side, unlike tomography, direct measurement requires only two bases. The modesty of this requirement can significantly simplify experiments. The simplification is particularly valuable in the context of quantum states whose dimension *d* is high: state tomography typically requires measurements of $O(d)$ bases. States with dimensions as high as a billion have been experimentally determined with generalisations of direct measurement [115, 124–126]. Many of these experiments benefited from numerous practical refinements of direct measurement: Increases in weak measurements' low SNRs [127–131]; enhancements of the procedure's directness [132, 133], particularly for multipartite systems [134]; and simplifications of the procedure (e.g. an avoidance of the scanning of the position measurement) [125, 135, 136]. These improvements have enabled further applications, as we now discuss.

Like quantum tomography, direct measurement has been generalised to fully characterise detectors and processes. Direct measurement was used to determine the projection effected by each outcome of a detector [137]. The number of bases measured is important for quantum-process determination. The number scales as $O(d^2)$ in standard process tomography: in different trials, one must input each element of each of the *d* bases and, for each state, measure all *d* bases. Direct measurement allows one to circumvent this rigmarole. As few as two measurements are required for each process parameter [113, 138, 139]. More recently, the ubiquitous Feynman propagator was observed via direct measurement [140]. Fundamentally, as for the wavefunction, direct measurement allows us to experimentally observe quantities that have been typically regarded as abstract mathematical elements of quantum physics.

5.3. Other direct-measurement procedures

Other ways of measuring the KD distribution do not involve weak measurements [141–146]. An early characterisation of a photon's transverse wavefunction relied on balanced homodyne detection [141]. This procedure determined the state with the use of a reference system (a local oscillator) in a superposition of position eigenstates [141]. The interference signal was proportional to the KD distribution.

Another direct method is based on optimal quantum cloning. After being proposed in [142], the method was experimentally demonstrated in [144]. According to this method, imperfect copies of the unknown quantum state are produced. These copies are optimal-universal-symmetric quantum clones [142], created via a controlled SWAP gate. See figure 7 for a distinct but related scheme. After this optimal cloning, one measures one copy's position while measuring the other copy's momentum. The distribution over the

Figure 7. Quantum circuit for measuring the KD distribution. $\hat{H}$ denotes the Hadamard gate, and $\hat{P} = \mathrm{diag}(1, i)$. The variable $s = 0, 1$ determines whether $\hat{P} = \mathrm{diag}(1, i)$ is implemented. The central unitary gate is a conditioned version of $\hat{U}_{A,\dots,Y}$, which cycles the quantum states such that $|\psi\rangle|\phi_A\rangle_A|\phi_B\rangle_B \cdots |\phi_Y\rangle_Y \rightarrow |\phi_Y\rangle|\psi\rangle_A|\phi_A\rangle_B \cdots |\phi_X\rangle_Y$. (For a k-extended KD distribution, $\hat{U}_{A,\dots,Y}$ can be implemented with $\mathcal{O}(k)$ conditioned SWAP gates.) If $s = 0$, the probability of finding the top qubit in $|0\rangle$ yields the KD quasi-probability's real part: $P_0^{s=0} = [1 + \mathrm{Re}\,(\langle y_l|x_k\rangle \dots \langle b_j|a_i\rangle \langle a_i|\hat{\rho}|y_l\rangle)]/2$. Similarly, if $s = 1$, the imaginary part can be measured: $P_0^{s=1} = [1 + \mathrm{Im}\,(\langle y_l|x_k\rangle \dots \langle b_j|a_i\rangle \langle a_i|\hat{\rho}|y_l\rangle)]/2$. This protocol was developed by Wagner *et al* [146] (see also [147–149]). The protocol is closely related to, but more general than, an earlier protocol [142], whose experimental implementation is reported on in [108]. The original protocol features a controlled SWAP gate used for optimal-universal-symmetric quantum cloning (see main text for details).

possible joint outcomes gives the KD distribution's real part: $\mathbb{P}(x, p; \hat{\rho}) = \mathrm{Re}\big(Q(x, p; \hat{\rho})\big)$. If a phase in the cloning process is changed, the protocol gives the imaginary part: $\mathbb{P}(x, p; \hat{\rho}) = \mathrm{Im}\big(Q(x, p; \hat{\rho})\big)$.

The three direct-measurement methods—the weak-measurement method and these latter two—can be seen in a common light. In each method, every trial involves measurements that project, to some degree, onto a superposition of position and momentum eigenstates. The projection onto $|x, p\rangle \equiv |x\rangle + \exp(i\phi)\,\langle p|x\rangle\,|p\rangle$ occurs with probability $\mathbb{P}(x, p) = \langle x, p|\hat{\rho}|x, p\rangle = \langle x|\hat{\rho}|x\rangle + \langle p|\hat{\rho}|p\rangle + \mathrm{Re}\big(\exp(i\phi)Q(x, p; \hat{\rho})\big)$. By varying the phase ϕ, one finds the KD distribution's real and imaginary parts.

We have described many ways in which direct measurement and its generalisations have been applied. Through direct measurements, one can observe elements of quantum physics, including wavefunctions, mixed states, processes and detection. In a practical sense, direct measurement simplifies the characterisation of these elements. In a fundamental sense, direct measurement allows us to observe these elements in their barest form as ensemble averages, sidestepping standard quantum tomography's shadows and inference.

6. The KD distribution in quantum thermodynamics

The KD distribution has benefited the modern quantum formulation of thermodynamics [31, 150–163]. Below, we review the treatment of classical thermodynamic work as a stochastic quantity. We then show how describing quantum work probabilistically produces inconsistencies. Conversely, the KD distribution satisfies reasonable requirements for a distribution describing work and heat fluctuations. Furthermore, a KD distribution's non-positivity signals contextuality—provable physical non-classicality. Although we lack space for the following topics, the KD distribution has found thermodynamic applications also in a consistent-histories framework for work fluctuations [151], an extension of quantum thermodynamics to quantum field theories [155], full-counting statistics [150, 156] and quadratic fermionic models [161].

The first and second laws of thermodynamics concern work, heat, and entropy. These quantities, when exchanged by small systems, fluctuate by amounts comparable to their averages. Fluctuation theorems are breakthroughs in non-equilibrium statistical mechanics (e.g. [164, 165]). In a prototypical example, a classical system has a Hamiltonian $H(\lambda_t, x, p)$ dependent on a time-varying control parameter λ_t and on a phase-space point (x, p). In a 'forward protocol', the system is prepared at a time $t = 0$ in a canonical ensemble $\{\exp\big(-\beta H(\lambda_0, x)\big)/Z(0)\}$ at an inverse temperature $\beta = 1/k_B T$. The partition function $Z(t) \equiv \int \mathrm{d}x \, \exp\big(-\beta H(\lambda_t, x)\big)$ normalises the distribution. The parameter is then ramped along an arbitrary trajectory λ_t arbitrarily quickly: $\lambda_0 \rightarrow \lambda_\tau$. In each trial of the protocol, the system absorbs a random amount W of work. After many trials, one can infer a probability distribution $\mathbb{P}_F^{cl}(W)$ over the possible W values. Similarly, define a 'reverse protocol' in which the canonical ensemble $\{\exp\big(-\beta H(\lambda_\tau, x)\big)/Z(\tau)\}$ is prepared at $t = 0$. The parameter is ramped as $\lambda_\tau \mapsto \lambda_0$ along the trajectory $\lambda_{\tau-t}$. One can infer a reverse work distribution $\mathbb{P}_R^{(cl)}(W)$. Each canonical ensemble has a free energy $-\frac{1}{\beta}\ln\big(Z(t)\big)$. Define the difference
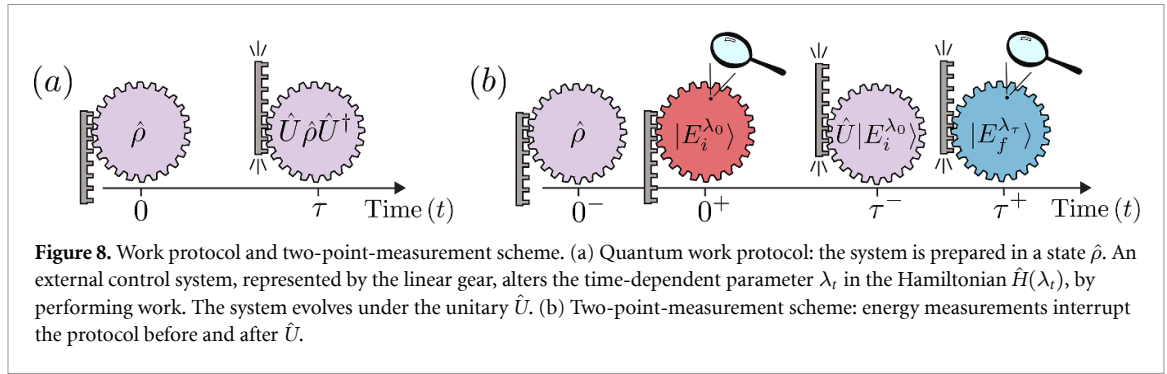
Figure 8. Work protocol and two-point-measurement scheme. (a) Quantum work protocol: the system is prepared in a state $\hat{\rho}$. An external control system, represented by the linear gear, alters the time-dependent parameter λ_t in the Hamiltonian $\hat{H}(\lambda_t)$, by performing work. The system evolves under the unitary $\hat{U}$. (b) Two-point-measurement scheme: energy measurements interrupt the protocol before and after $\hat{U}$.

$\Delta F \equiv -\frac{1}{\beta}\ln\left(Z(\tau)/Z(0)\right)$ between the initial and final Hamiltonians' equilibrium free energies. In terms of ΔF, *Crooks' theorem* reveals a symmetry of the two distributions [165]:

$$\frac{\mathbb{P}_F^{cl}(W)}{\mathbb{P}_R^{cl}(-W)} = e^{\beta(W-\Delta F)}. \tag{61}$$

To understand this equality intuitively, consider a large work value $W \gg \Delta F$. The exponential $e^{\beta(W-\Delta F)}$ is enormous. Hence $\mathbb{P}_F^{cl}(W) \gg \mathbb{P}_R^{cl}(-W)$: one is far more likely to pay a large amount W of work, during a forward trial, than to recoup an amount W during a reverse trial. Crooks' theorem implies *Jarzynski's equality*,

$$\langle e^{-\beta W} \rangle_F = e^{-\beta\Delta F}, \tag{62}$$

which implies the second-law-like inequality $\langle W \rangle_F \geqslant \Delta F$ [164, 166]. [The averages are with respect to $\mathbb{P}_F^{cl}(W)$.] Hence Crooks' theorem and Jarzynski's equality are said to strengthen the second law. They are also called *fluctuation theorems*. Using either, one can infer ΔF from out-of-equilibrium experiments [167]. This fact is useful, as ΔF is difficult to measure [168].

Measurement disturbance complicates a quantum extension of this protocol. One approach, the *two-point-measurement scheme*, involves the following forward protocol (figure 8). Consider a quantum system prepared in a canonical state $\hat{\rho} \propto \exp\left(-\beta\hat{H}(\lambda_0)\right)$ and evolved under a quantum Hamiltonian $\hat{H}(\lambda_t)$ dependent on a parameter λ_t. $\hat{H}(\lambda_t)$ eigen-decomposes as $\hat{H}(\lambda_t) = \sum_j E_j(\lambda_t)\hat{\Pi}_j(\lambda_t)$, with eigenvalues $E_j(\lambda_t)$ and eigen-projectors $\hat{\Pi}_j(\lambda_t)$. At time $t = 0$, one measures the energy, obtaining an outcome $E_j(\lambda_0)$ and projecting the state onto $\hat{\Pi}_j(\lambda_0)/\mathrm{Tr}\left(\hat{\Pi}_j(\lambda_0)\right)$. From $t = 0$ to $t = \tau$, the unitary $\hat{U} = \mathcal{T}e^{-i\int_0^\tau \hat{H}(\lambda_t)dt/\hbar}$ evolves the system. $\mathcal{T}$ denotes the time-ordering operation. At $t = \tau$, one measures the energy, obtaining an outcome $E_k(\lambda_\tau)$ and projecting the state onto $\hat{\Pi}_k(\lambda_\tau)/\mathrm{Tr}\left(\hat{\Pi}_k(\lambda_\tau)\right)$. The joint probability of obtaining outcomes $E_j(\lambda_0)$ and $E_k(\lambda_\tau)$ is

$$P_{j,k}(\hat{\rho}) \equiv \mathrm{Tr}\left(\hat{U}^\dagger \hat{\Pi}_k(\lambda_\tau)\hat{U}\hat{\Pi}_j(\lambda_0)\hat{\rho}\hat{\Pi}_j(\lambda_0)\right). \tag{63}$$

The probability density associated with spending an amount W of work during any given forward trial is

$$\mathbb{P}_F(W) = \sum_{j,k} P_{j,k}(\hat{\rho})\,\delta\left(W - \left[E_k(\lambda_\tau) - E_j(\lambda_0)\right]\right). \tag{64}$$

One can define $\mathbb{P}_R(W)$ similarly for a reverse protocol, then prove Crooks' theorem.

This two-point-measurement scheme faces a problem if $\hat{\rho}$ has coherence with respect to the initial energy eigenbasis [the $\hat{H}(\lambda_0)$ eigenbasis]. The initial measurement dephases $\hat{\rho}$, affecting the marginal probability $\sum_j P_{j,k}(\hat{\rho})$. One might expect this marginal to equal the probability $\mathrm{Tr}\left(\hat{\Pi}_k(\lambda_\tau)\hat{U}\hat{\rho}\hat{U}^\dagger\right)$ of obtaining the final-measurement outcome $E_k(\lambda_\tau)$, in the absence of any initial measurement. Yet the two probabilities do not equal each other: $\sum_j P_{j,k}(\hat{\rho}) \neq \mathrm{Tr}\left(\hat{\Pi}_k(\lambda_\tau)\hat{U}\hat{\rho}\hat{U}^\dagger\right)$. This lack of equality stems from measurement disturbance, caused by the mutual non-commutativity of $\hat{\rho}$, $\hat{H}(\lambda_0)$ and $\hat{H}(\lambda_\tau)$. Nevertheless, a KD distribution—a set of quasi-probabilities—has the desired marginal $\mathrm{Tr}\left(\hat{\Pi}_k(\lambda_\tau)\hat{U}\hat{\rho}\hat{U}^\dagger\right)$, as well as the similarly desired marginal $\mathrm{Tr}\left(\hat{\Pi}_j(\lambda_0)\hat{\rho}\right)$:

$$Q_{j,k}(\hat{\rho}) \equiv \mathrm{Tr}\left(\hat{U}^\dagger \hat{\Pi}_k(\lambda_\tau)\hat{U}\hat{\Pi}_j(\lambda_0)\hat{\rho}\right). \tag{65}$$

In terms of this joint KD quasi-probability, one can define the KD work distribution [31]

$$\tilde{\mathbb{P}}_{\text{F}}(W) \equiv \sum_{j,k} Q_{j,k}(\hat{\rho})\, \delta\left(W - \left[E_k(\lambda_\tau) - E_j(\lambda_0)\right]\right). \tag{66}$$

A Jarzynski-like equality (section 7) depends on an extended KD distribution over analogues of work [30].

The two-point-measurement joint distribution (63) and KD joint distribution (65) share several properties. We recover the marginal $P_{j,k}(\hat{\rho})$ from $Q_{j,k}(\hat{\rho})$ if $\hat{\rho}$ is block-diagonal with respect to $\hat{H}(\lambda_0)$'s eigenbasis. Therefore, $Q_{j,k}(\hat{\rho})$ leads to fluctuation theorems for such states $\hat{\rho}$. We recover $P_{j,k}(\hat{\rho})$ from $Q_{j,k}(\hat{\rho})$ also if one measures coarse-grained observables [152]. As shown in section 2.2, $Q_{j,k}(\hat{\rho})$ has marginals equal to the Born-rule predictions about preparing $\hat{\rho}$ and evolving the state under $\hat{U}$:

$$\sum_j Q_{j,k}(\hat{\rho}) = \text{Tr}\left(\hat{\Pi}_k(\lambda_\tau)\,\hat{U}\hat{\rho}\hat{U}^\dagger\right), \quad \sum_k Q_{j,k}(\hat{\rho}) = \text{Tr}\left(\hat{\Pi}_j(\lambda_0)\,\hat{\rho}\right) \quad \text{and} \quad \sum_{j,k} Q_{j,k}(\hat{\rho}) = 1. \tag{67}$$

Additionally, a work distribution should lead to an average work value equal to the usual definition $\langle W \rangle \equiv \text{Tr}\left(\hat{H}(\lambda_\tau)\,\hat{U}\hat{\rho}\hat{U}^\dagger\right) - \text{Tr}\left(\hat{H}(\lambda_0)\,\hat{\rho}\right)$. The two-point-measurement distribution (equation (64)) does not always satisfy this condition: $\int \mathbb{P}_{\text{F}}(W)\,\mathrm{d}W \neq \langle W \rangle$, generally. Yet the KD work distribution does, due to equations (67): $\int \tilde{\mathbb{P}}_{\text{F}}(W)\,\mathrm{d}W = \langle W \rangle$. The KD average differs from the two-point-measurement average, and the KD variance differs from the two-point-measurement variance, whenever $\hat{\rho}$ has energy coherences [159]. Furthermore, when $[\hat{U}^\dagger \hat{H}(\lambda_\tau)\hat{U}, \hat{H}(\lambda_0)] \neq 0$, the KD work variance can assume values only from a subset of the values achievable if work is a classical random variable [154].

No-go theorems demonstrate the incompatibility of reasonable criteria for work probability distributions $\mathbb{P}(W)$ [154, 158, 169]. One criterion is convex-linearity in $\hat{\rho}$ [153, 154, 169], a property that $Q_{j,k}(\hat{\rho})$ satisfies. Yet, if $[\hat{U}^\dagger \hat{H}(\lambda_\tau)\hat{U}, \hat{H}(\lambda_0)] \neq 0$, one cannot define a joint probability with convex-linearity and the marginal property (67) [158]. Another no-go theorem replaces the marginal property with two alternative criteria [169]: (i) $\int \mathbb{P}(W)\,\mathrm{d}W$ equals the average energy change $\langle W \rangle$. (ii) If $\hat{\rho}$ is block-diagonal with respect $\hat{H}(\lambda_0)$'s eigenbasis, then $\mathbb{P}(W) = \mathbb{P}_{\text{F}}(W)$. No probability distribution satisfies the convex-linearity criterion in addition to criteria (i) and (ii) [169]. However, the KD work distribution $\tilde{\mathbb{P}}_{\text{F}}(W)$ is convex-linear in $\hat{\rho}$ and satisfies the marginal property and criteria (i) and (ii), as well as other reasonable criteria [162]. Alternative work-distribution criteria are satisfied only by a Terletsky–Margenau–Hill distribution (a KD distribution's real part) [157, 162].
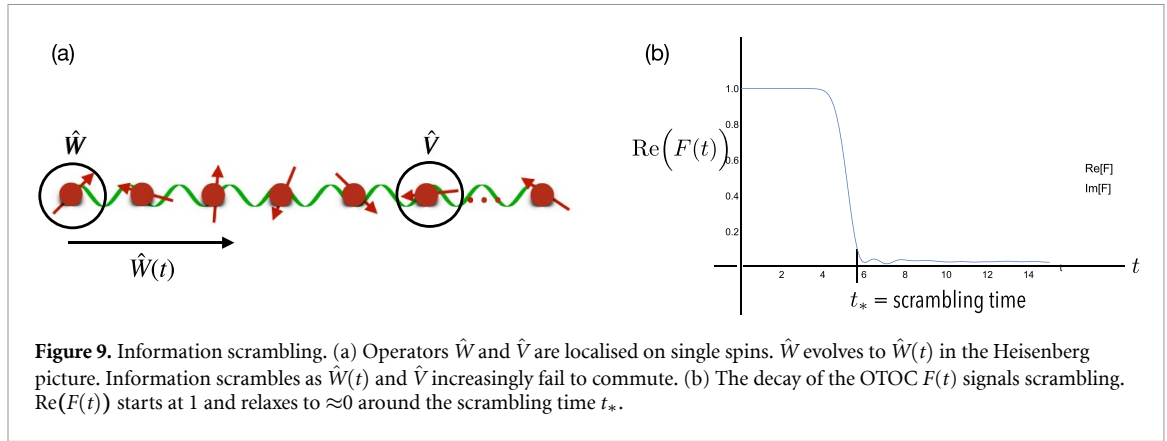
In addition to satisfying the criteria above, KD distributions signal non-classicality in a quantum engine's linear-response work output. *Linear-response theory* describes how systems react to small perturbations. Consider a quantum system subject to a static Hamiltonian $\hat{H}_0$ perturbed by a time-dependent potential $g\hat{V}(t)$, where $g \ll 1$. Let the potential be cyclic: $\hat{V}(0) = \hat{V}(\tau) = 0$. In the interaction picture, the potential has the form $\hat{V}_I(t) = \mathrm{e}^{\mathrm{i}\hat{H}_0 t/\hbar}\hat{V}(t)\mathrm{e}^{-\mathrm{i}\hat{H}_0 t/\hbar}$ and the time average $\bar{\hat{V}}_I(\tau) = \frac{1}{\tau}\int_0^\tau \hat{V}_I(t)\,\mathrm{d}t$. The system is prepared in $\hat{\rho}$ and evolved unitarily under $\hat{H}(t) = \hat{H}_0 + g\hat{V}(t)$. On average, the system produces an amount of work [170]

$$\langle W \rangle = \frac{2g\tau}{\hbar}\text{Im}\left(\text{Tr}\left(\hat{H}_0\bar{\hat{V}}_I(\tau)\hat{\rho}\right)\right) + O(g^2). \tag{68}$$

If $\text{Im}(\text{Tr}(\hat{H}_0\bar{\hat{V}}_I(\tau)\hat{\rho})) \neq 0$, then the average work is first-order in the perturbation strength: $\langle W \rangle = O(g)$. The lack of equality translates into a condition under which a KD distribution signals contextuality (rigorous non-classicality), as detailed in section 8.3.2. In another work-extraction setting, KD non-positivity signals that non-commutation can enhance the achievable power [171].

Complementary to work expenditure is heat exchange. Consider heat baths A and B, with respective Hamiltonians $\hat{H}^{\text{A}}$ and $\hat{H}^{\text{B}}$, isolated from their environment. Let the initial state $\hat{\rho}^{\text{AB}}$ be a tensor product of thermal states with inverse temperature β^{A} and $\beta^{\text{B}} < \beta^{\text{A}}$: $\hat{\rho}^{\text{AB}} \propto \exp(-\beta^{\text{A}}\hat{H}^{\text{A}}) \otimes \exp(-\beta^{\text{B}}\hat{H}^{\text{B}})$. Consider any unitary $\hat{U}$ that conserves the global energy: $[\hat{U}, \hat{H}^{\text{A}} + \hat{H}^{\text{B}}] = 0$. The second law of thermodynamics implies that heat flows from the hotter B to the colder A on average. The net heat $\mathcal{Q}$ transferred from B to A satisfies an *exchange* fluctuation theorem [172]: $\mathbb{P}_{\text{F}}(+\mathcal{Q})/\mathbb{P}_{\text{R}}(-\mathcal{Q}) = \exp([\beta^{\text{B}} - \beta^{\text{A}}]\mathcal{Q})$. This equation governs classical and quantum baths alike, if we define $\mathcal{Q}$ through projective energy measurements.

Non-classical heat flows may occur if $\hat{\rho}^{\text{AB}}$ has thermal local marginals, $\text{Tr}_{\text{B}}(\hat{\rho}^{\text{AB}}) = \exp(-\beta^{\text{A}}\hat{H}^{\text{A}})/Z^{\text{A}}$ and $\text{Tr}_{\text{A}}(\hat{\rho}^{\text{AB}}) = \exp(-\beta^{\text{B}}\hat{H}^{\text{B}})/Z^{\text{B}}$, but is correlated. The correlations can drive heat from the colder A to the hotter B on average. Quantum correlations can do so at a rate unachievable classically: Let A and B have d-dimensional Hilbert spaces. An amount of heat $\mathcal{Q} < \log(d)/(\beta^{\text{B}} - \beta^{\text{A}})$ constitutes an *anomalous backflow*, requiring entanglement [173, 174]. The required correlations disappear under strong energy measurements, à la the two-point-measurement scheme. Measuring weakly preserves the anomalous backflow and suggests a KD heat quasi-probability distribution [152]. Negative and non-real KD quasi-probabilities signal

Figure 9. Information scrambling. (a) Operators $\hat{W}$ and $\hat{V}$ are localised on single spins. $\hat{W}$ evolves to $\hat{W}(t)$ in the Heisenberg picture. Information scrambles as $\hat{W}(t)$ and $\hat{V}$ increasingly fail to commute. (b) The decay of the OTOC $F(t)$ signals scrambling. $\mathrm{Re}(F(t))$ starts at 1 and relaxes to ≈ 0 around the scrambling time t_*.

non-classical heat [152] and work flows [158]. Aside from energy, A and B can exchange other quantities, such as particles [175]. The quantities need not commute with each other, as exemplified by the x-, y-, and z-components of spin [176]. An extended KD distribution characterises such non-commuting quantities' fluctuations, which can signal contextuality [160].

7. The KD distribution behind the out-of-time-ordered correlator (OTOC)

The *OTOC* [30] is a witness of many-body quantum chaos. In addition to elucidating superconductors [177] and black holes [178–181], the OTOC has been observed with nuclear magnetic resonance [182–185], trapped ions [186–190] and other platforms [191–204]. The OTOC equals an average over an extended KD distribution [30, 45]. We present the distribution after reviewing the OTOC. The distribution signals a fluctuation-type theorem—resembling an extension of the second law of thermodynamics (section 6)—that contains the OTOC [30]. The fluctuation-type theorem suggested new techniques for measuring the OTOC experimentally [30, 45, 205, 206]. Also, the extended KD distribution distinguishes between chaos and decoherence [40]. Further applications lie beyond the scope of this review: The extended KD distribution diagnoses a quantum kicked top's chaos [207] and features in an uncertainty relation for quantum chaos [208].

The OTOC signals *scrambling*, the spreading of initially localised information through many-body entanglement [209]. Consider a quantum system of $N \gg 1$ degrees of freedom. Let a non-local, non-integrable ('chaotic') Hamiltonian $\hat{H}$ govern the system, which is in a state $\hat{\rho}$. Denote expectation values by $\langle \bullet \rangle \equiv \mathrm{Tr}(\bullet \hat{\rho})$. Figure 9 illustrates scrambling with an N-qubit chain in a thermal state $\hat{\rho}_\beta \equiv \exp(-\beta \hat{H})/\mathcal{Z}$. The inverse temperature is β, and the partition function $\mathcal{Z}$ normalises the state. Denote by $\hat{W}$ and $\hat{V}$ unitary or Hermitian operators localised far apart[12]. In our example, we denote by $\hat{\sigma}_a^{(j)}$ component $a = x, y, z$ of qubit j's spin. The spin chain can have $\hat{W} = \hat{\sigma}_x^{(1)}$ and $\hat{V} = \hat{\sigma}_z^{(N)}$. $\hat{W}$ could be a perturbation, whereas $\hat{V}$ could be an observable measured later, in an attempt to recover information about the perturbation. The operators eigendecompose as $\hat{W} = \sum_{w_j} w_j \hat{\Pi}_{w_j}^W$ and $\hat{V} = \sum_{v_j} v_j \hat{\Pi}_{v_j}^V$. Under the unitary $\hat{U} = \exp(-i\hat{H}t/\hbar)$, $\hat{W}$ evolves to $\hat{W}(t) \equiv \hat{U}^\dagger \hat{W} \hat{U}$ in the Heisenberg picture. The more $\hat{W}(t)$ spreads, the more $[\hat{W}(t), \hat{V}]$ grows. Yet the simple correlator $\langle [\hat{W}(t), \hat{V}] \rangle$ can vanish at late times, due to cancellations between $\hat{\rho}$ eigenstates in the average. The average $C(t) \equiv \langle |[\hat{W}(t), \hat{V}]|^2 \rangle$, being of a non-negative squared magnitude, avoids this pitfall. $C(t)$ more reliably certifies scrambling—the spreading of the information initially localised in $\hat{W}$. If $\hat{W}$ and $\hat{V}$ are unitary, $C(t)$ decomposes as $C(t) = 2[1 - 2\mathrm{Re}(F(t))]$. The OTOC therein is defined as

$$F(t) \equiv \langle \hat{W}^\dagger(t) \hat{V}^\dagger \hat{W}(t) \hat{V} \rangle \equiv \mathrm{Tr}\left(\hat{W}^\dagger(t) \hat{V}^\dagger \hat{W}(t) \hat{V} \hat{\rho} \right). \tag{69}$$

A decay of $F(t)$ to ≈ 0 signals scrambling, illustrated in figure 9(a). The OTOC decays exponentially at early times under highly non-local interactions: $\mathrm{Re}(F(t)) \sim 1 - \frac{1}{N} \exp(\lambda_\mathrm{L} t)$. λ_L resembles a Lyapunov exponent, which controls how phase-space trajectories diverge under classical chaos [210].

Halpern *et al* decomposed the OTOC in terms of an extended KD distribution [30, 45],

$$Q_{v_1, w_1, v_2, w_2}(\hat{\rho}(t)) = \mathrm{Tr}\left(\hat{\Pi}_{w_2}^{W(t)} \hat{\Pi}_{v_2}^V \hat{\Pi}_{w_1}^{W(t)} \hat{\Pi}_{v_1}^V \hat{\rho} \right). \tag{70}$$

---

[12] Some OTOC results require $\hat{W}$ and $\hat{V}$ to be Hermitian; some, for the operators to be unitary; and some, for the operators to be both. The quasi-probability results highlighted here are general, holding if $\hat{W}$ and $\hat{V}$ are merely normal and thus diagonalisable. That is, $\hat{W}$ and $\hat{V}$ can be Hermitian or unitary.

In illustrative qubit-chain examples, $Q(\hat{\rho}(t))$ typically has a negative real component and a small imaginary component [45]. The OTOC equals an average over $Q(\hat{\rho}(t))$ [45]:

$$F(t) = \sum_{v_1,w_1,v_2,w_2} v_1 w_1 v_2^* w_2^* \, Q_{v_1,w_1,v_2,w_2}(\hat{\rho}(t)) \equiv \langle v_1 w_1 v_2^* w_2^* \rangle_{Q(\hat{\rho}(t))}. \tag{71}$$

In the spin-chain example, since W and V are Pauli operators, each sum runs over the eigenvalues ± 1. The OTOC's decomposition motivated the 21st-century rediscovery of extended KD distributions [30, 45]. More thermodynamically, the OTOC decomposes also in terms of the distribution's characteristic function, or the Fourier transform of the distribution: $\langle e^{\beta_1 v_1 + \beta_1' w_1 + \beta_2 v_2^* + \beta_2' w_2^*} \rangle_{Q(\hat{\rho}(t))}$.[13] More precisely, the OTOC follows from differentiating the characteristic function:

$$F(t) = \frac{\partial^4}{\partial \beta_1 \, \partial \beta_1' \, \partial \beta_2 \, \partial \beta_2'} \left\langle e^{\beta_1 v_1 + \beta_1' w_1 + \beta_2 v_2^* + \beta_2' w_2^*} \right\rangle_{Q(\hat{\rho}(t))} \bigg|_{\beta_1,\beta_1',\beta_2,\beta_2'=0}, \quad \beta_1, \beta_1', \beta_2, \beta_2' \in \mathbb{R}. \tag{72}$$

Equation (72) resembles Jarzynski's equality (equation (62) in section 6), an extension of the second law of thermodynamics, in two ways. First, equation (72) and Jarzynski's equality suggest schemes for extracting difficult-to-measure quantities from more-easily-measurable quantities: Using Jarzynski's equality, we can infer a free-energy difference ΔF from a probability distribution over the possible amounts of the work spent on a non-equilibrium protocol [164, 211]. Similarly, the Jarzynski-like equality (72) enables new methods for measuring $F(t)$ experimentally [30]: Equation (69) shows that the OTOC is neither a probability nor an expectation value. How to measure the OTOC was therefore unclear for a while. The initially proposed measurement schemes rely on interference [212], Ramsey interferometry [213] and a quantum clock [214]. Linking the OTOC to $Q(\hat{\rho}(t))$ unlocked new toolkits for measuring the OTOC [30, 45]. First, the OTOC can be inferred from weak measurements [30, 45]. Second, if $\hat{W}(t)$ and $\hat{V}$ square to the identity, $\hat{W}(t)^2 = \hat{V}^2 = \mathbb{1}$, then one can infer the OTOC from arbitrary-strength measurements [205, 206]. Pauli operators illustrate this property: $\left(\hat{\sigma}_a^{(j)}\right)^2 = \hat{\mathbb{1}}$.

The second way in which equation (72) resembles Jarzynski's equality is, both equations cast the object of interest [ΔF or $F(t)$] in terms of a moment-generating function. Beyond these two parallels, Equation (72) inspired a fluctuation theorem for the spreading of correlations from a system into an environment [215].

A numerical study established $Q(\hat{\rho}(t))$ as a scrambling witness more robust than $F(t)$ [40]. An experimentalist may measure $F(t)$, observe a decay, and conclude that their system likely contains many-body entanglement. Yet this decay may stem from decoherence, rather than scrambling [201, 216–218]. If used to detect scrambling, therefore, $F(t)$ can lead to false positives. $Q(\hat{\rho}(t))$ overcomes this pitfall, distinguishing scrambling from integrable Hamiltonians despite decoherence [40]. Alonso *et al* simulated $N = 5$ superconducting qubits undergoing Lindblad dynamics. The qubits evolve under a transverse-field Ising model with or without a longitudinal field:
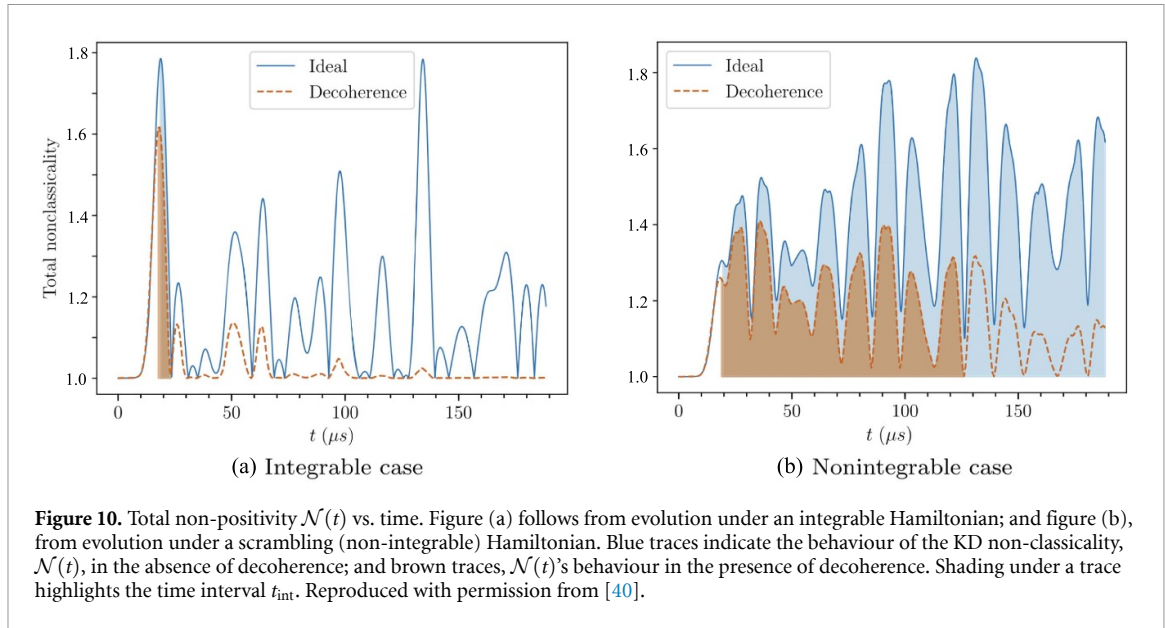
$$\hat{H} = -J \sum_{j=1}^{N-1} \hat{\sigma}_z^{(j)} \hat{\sigma}_z^{(j+1)} - g \sum_{j=1}^{N} \hat{\sigma}_x^{(j)} - h \sum_{j=1}^{N} \hat{\sigma}_z^{(j)}. \tag{73}$$

In the absence of the longitudinal field, if $(h/J, g/J) = (0.000, 1.05)$, $\hat{H}$ is integrable. In the presence of the longitudinal field, if $(h/J, g/J) = (0.500, 1.05)$, $\hat{H}$ is scrambling (non-integrable). The local operators used were $\hat{W} = \hat{\sigma}_1^z$ and $\hat{V} = \hat{\sigma}_N^z$. Figure 10 shows the quasi-probability non-classicality, $\mathcal{N}(t) \equiv \sum |Q(\hat{\rho}(t))|$ (section 9.3), plotted against time. $\mathcal{N}(t)$ begins at 1, grows, peaks, returns to 1, and then keeps oscillating. Denote by t_{int} the time interval from $\mathcal{N}(t)$'s first peak to the next 1. t_{int} distinguishes scrambling from integrable behaviour: in the presence of decoherence, t_{int} is an order of magnitude longer under scrambling than under integrable dynamics. $Q(\hat{\rho}(t))$ thus witnesses scrambling more robustly than $F(t)$ does.

8. The KD distribution and the foundations of quantum mechanics

One way to define the notion of a classical experiment is to say that joint probability distributions describe the corresponding system's preparation, manipulation and measurement [2–12]. Under this definition, an experiment has been proven to be classical if, and only if, it is non-contextual [11] (we define this term below). Thus, if every quasi-probabilistic description of an experiment is non-positive, the experiment is

---

[13] This expression technically contains a Laplace transform, rather than a Fourier transform, since the βs are real numbers, rather than imaginary. However, the fluctuation-theorem community calls such transforms 'Fourier'.

Figure 10. Total non-positivity $\mathcal{N}(t)$ vs. time. Figure (a) follows from evolution under an integrable Hamiltonian; and figure (b), from evolution under a scrambling (non-integrable) Hamiltonian. Blue traces indicate the behaviour of the KD non-classicality, $\mathcal{N}(t)$, in the absence of decoherence; and brown traces, $\mathcal{N}(t)$'s behaviour in the presence of decoherence. Shading under a trace highlights the time interval t_{int}. Reproduced with permission from [40].

non-classical. Every experiment admits of infinitely many quasi-probability representations, however. Therefore, the existence of a single non-positive quasi-probability representation does not allow one to conclude that the experiment is non-classical. Nevertheless, operationally defined KD distributions' non-positivity has been linked to non-classical physics in several settings. In this section, we outline several examples.

8.1. Generalised Leggett–Garg inequality

Leggett and Garg designed a Gedanken experiment for testing the limits of macroscopic coherence and notions of 'quantumness' [219]. In the spirit of Bell, Leggett and Garg formulated an inequality for data acquired from sequential measurements. The inequality is satisfied under the assumptions of macroscopic realism and non-invasive detectability. However, quantum experiments generally can violate the inequality. Thus, the Leggett–Garg inequality has become a popular tool for distinguishing macro-realistic theories from quantum mechanics. Williams and Jordan [220] have connected anomalous weak values [78] with violations of the weak-measurement version of the Leggett–Garg inequality [221–223]. (For details about weak values and weak measurements, see section 4.) Such violations have been observed experimentally [224–227]. We now elaborate on this connection.

In experimental tests of the weak-measurement version of the Leggett–Garg inequality, three measurements are performed: a strong, followed by a weak, followed by another strong measurement. This set-up allows one to verify quantum behaviour using one experimental configuration, rather than the three or four configurations needed for the original Leggett–Garg approach [219]. The three outcomes are $r^{(1)}$, $r^{(2)}$ and $r^{(3)}$. One analyses the correlation functions $K_{ij} = \langle r^{(i)} r^{(j)} \rangle$. The superscripts label when a measurement takes place. The measurement outcomes are scaled such that the statistical average $\langle r_1 \rangle = \langle \psi | \hat{A} | \psi \rangle$, where $\hat{A}$ denotes a measured observable and $|\psi\rangle$ denotes the prepared quantum state. Analogous statements concern r_2 and r_3, together with observables $\hat{B}$ and $\hat{C}$.

The simplest case involves a two-level system. The three observables are $\hat{A} = |a_1\rangle\langle a_1| - |a_2\rangle\langle a_2|$, $\hat{B} = |b_1\rangle\langle b_1| - |b_2\rangle\langle b_2|$ and $\hat{C} = |c_1\rangle\langle c_1| - |c_2\rangle\langle c_2|$. The observables' eigenvalues are $a_1 = b_1 = c_1 = +1$ and $a_2 = b_2 = c_2 = -1$. The generalised Leggett–Garg inequality is

$$\mathcal{L} \equiv K_{12} + K_{23} - K_{13} \leqslant 1. \tag{74}$$

One can achieve the classical upper bound if the first and the second measurements are completely correlated, as are the second and third: $K_{12} = K_{23} = 1$. Classically, also the first and the third measurements would consequently be completely correlated: $K_{13} = 1$, so $\mathcal{L} = 1$. Classically, no other configuration can supersede this value of $\mathcal{L} = 1$.

When the intermediate measurement is weak, we can express $\mathcal{L}$ in terms of the KD distribution $Q_{j,k}(|a_i\rangle\langle a_i|) \equiv \langle a_i | c_k \rangle \langle c_k | b_j \rangle \langle b_j | a_i \rangle$. For simplicity, we suppose that the quantum system begins in an $\hat{A}$ eigenstate $|a_i\rangle$. The first strong measurement yields the eigenvalue a_i. It is followed by a weak measurement of $\hat{B}$ and a strong measurement of $\hat{C}$. In this weak-measurement limit, we can replace the correlation

function with the approximation

$$\mathcal{L} \approx \mathrm{Re}\left(\langle a_i | \hat{A}\hat{B} + \hat{B}\hat{C} - \hat{A}\hat{C} | a_i \rangle\right). \tag{75}$$

This is the form of the correlation function considered by Leggett and Garg [219]. We stress that the $\hat{B}$ measurement's weakness allows us to neglect that measurement's influence on the $\hat{A}\hat{C}$ correlator. If the intermediate measurement is of finite strength, this assumption breaks down. Williams and Jordan [220] describes the measurement's influence on the corrrelator. Let us insert the operators' spectral expansions into equation (75). We obtain an expression dependent on the KD distribution:

$$\mathcal{L} \approx \sum_{j,k} \mathrm{Re}\left(Q_{j,k}\left(|a_i\rangle\langle a_i|\right)\right)\left(a_i b_j + b_j c_k - a_i c_k\right). \tag{76}$$

The rightmost parenthetical factor is upper-bounded by 1. Hence we see that, for the right-hand side to exceed the upper bound of 1, some joint KD quasi-probability $Q_{j,k}(|a_i\rangle\langle a_i|)$ must have a negative real part. This bound violation was observed experimentally. Furthermore, this connection between Leggett–Garg-type inequalities and quasi-probability negativity was observed also in [228, 229].

The Leggett–Garg inequality further underscores the connection between weak values and the KD distribution. Weakly measuring the intermediate observable $\hat{B}$, given strong pre- and post-selective measurements, is equivalent to measuring the weak values $B_{\mathrm{w}}(a_i, c_k) \equiv \frac{\langle c_k | \hat{B} | a_i \rangle}{\langle c_k | a_i \rangle}$ (equation (48)). Thus, in the weak measurement limit, $\mathcal{L}$ can be expressed as

$$\mathcal{L} \approx \sum_k P(c_k | a_i)\left[(a_i + c_k)\mathrm{Re}\left(B_{\mathrm{w}}\left(a_i, c_k\right)\right) - a_i c_k\right]. \tag{77}$$

The $P(c_k | a_i) = |\langle c_k | a_i \rangle|^2$ denotes the conditional probability that, if the first measurement yields a_i and the $\hat{B}$ measurement is weak, the final measurement yields c_k.

The right-hand side of equation (77) is a convex sum. Therefore, for the right-hand side to exceed the classical upper bound of 1, the factor in square brackets must exceed 1 for some a_i and c_k. When $\mathrm{Re}(B_{\mathrm{w}}) > 0$, this condition can be met only when $a_i = c_k = +1$ (when $i = k = 1$). Otherwise, the factor in curly brackets is negative. In such a case, the bound-violation condition simplifies to $2\mathrm{Re}(B_{\mathrm{w}}(a_1, c_1) - 1) > 1$, or $\mathrm{Re}(B_{\mathrm{w}}(a_1, c_1)) > 1$. That is, the weak value's real part must exceed the greatest $\hat{B}$ eigenvalue. If, instead, $\mathrm{Re}(B_{\mathrm{w}}) < 0$, the classical bound $\mathcal{L} \leqslant 1$ can be violated only when $a_i = c_k = -1$ (when $i = k = 2$). In this case, the condition necessary for violating the Leggett–Garg inequality simplifies to $-2\mathrm{Re}(B_{\mathrm{w}}(a_2, c_2) - 1) > 1$, or $\mathrm{Re}(B_{\mathrm{w}}(a_2, c_2)) < -1$. Again, the weak value must lie outside the $\hat{B}$ spectrum.

These relations establish a one-to-one correspondence between a Leggett–Garg-inequality violation and an anomalous weak value. Expanding the conditions in terms of KD quasi-probabilities yields $\sum_j b_j \mathrm{Re}(Q_{j,k=1}(|a_1\rangle\langle a_1|)) > \sum_j \mathrm{Re}(Q_{j,k=1}(|a_1\rangle\langle a_1|))$, implying $\mathrm{Re}(Q_{j=2,k=1}(|a_1\rangle\langle a_1|)) < 0$, or $\sum_j b_j Q_{j,k=2}(|a_2\rangle\langle a_2|) < -\sum_j \mathrm{Re}(Q_{j,k=2}(|a_2\rangle\langle a_2|))$, implying $\mathrm{Re}(Q_{j,k=2}(|a_2\rangle\langle a_2|)) < 0$. We can now interpret, in terms of KD quasi-probabilities, the conditions necessary for violating the Leggett–Garg inequality's classical upper bound of 1, as has been observed experimentally. At least one of two KD quasi-probabilities, $Q_{j=2,k=1}(|a_1\rangle\langle a_1|)$ or $Q_{j=1,k=2}(|a_2\rangle\langle a_2|)$, must have a negative real part. Each such negative real part causes a corresponding weak value, $\mathrm{Re}(B_{\mathrm{w}}(a_1, c_1))$ or $\mathrm{Re}(B_{\mathrm{w}}(a_2, c_2))$, to lie outside the $\hat{B}$ spectrum.

8.2. The KD distribution and the consistent-histories interpretation of quantum mechanics

The mathematical framework of quantum mechanics is well-established but not a complete physical theory. The role of an *interpretation* of quantum mechanics [230] is to append to the mathematical formalism a theory of real underlying (ontic) physics that results in the quantum phenomena we observe in laboratories. Such interpretations include the many-worlds interpretation [231, 232], de Broglie–Bohm mechanics [233] and the consistent-histories interpretation [9]. The latter has a deep connection with the KD distribution. We review this connection below.

The consistent-histories interpretation of quantum mechanics predicts the same measurement probabilities as textbook quantum mechanics. The Copenhagen interpretation [234], however, suggests that quantum states have well-defined properties only when measured. In contrast, the consistent-histories interpretation provides a framework that (sometimes) describes properties of quantum particles between observations. The interpretation was developed by several scholars [235–238], most notably by Griffiths [9, 239].

In the consistent-histories interpretation, properties (or physical attributes) of quantum states are defined in accordance with von Neumann's theory [9, 81]. A property is associated with a projective operator

onto a subspace of a Hilbert space[14]. For example, 'position' is not a property. However, 'the particle is at position x' is. Consider two sets of projective measurement operators, $\mathcal{A} = \{\hat{\Pi}_i^{(1)}\}$ and $\mathcal{B} = \{\hat{\Pi}_j^{(2)}\}$. Quantum theory struggles to describe quantum states' underlying because the question *does the state have properties $\hat{\Pi}_1^{(1)}$ and $\hat{\Pi}_3^{(2)}$?* seems difficult to answer when $[\hat{\Pi}_1^{(1)}, \hat{\Pi}_3^{(2)}] \neq 0$. Under certain consistency conditions, the consistent-histories interpretation facilitates a description of underlying physical properties of a quantum state's time-evolution through Hilbert space.

Next, we review the technical framework for the consistent-histories interpretation. Afterwards, we connect the consistency criterion to a KD distribution. Finally, we provide an example.

Consider a pure initial quantum state $\hat{\rho}_i$. Consider also a set of times $t_0 < t_1 < \ldots < t_k < t_f$. We work in the Heisenberg picture, in which quantum states are constant in time and observables evolve unitarily. Define k sets of projector-valued measures $\mathcal{A}^{(l)} = \{\hat{\Pi}_{i_l}^{(l)}(t_l)\}$, where $\sum_{i_l} \hat{\Pi}_{i_l}^{(l)}(t_l) = \hat{\mathbb{1}}$ and $l = 1, 2, \ldots, k$. We convert a Schrödinger-picture observable $\hat{X}$ to a Heisenberg-picture observable through $\hat{X}(t_i) = \hat{U}^\dagger(t_0, t_i) \hat{X} \hat{U}(t_0, t_i)$. Finally, consider a final rank-1 observable $\hat{\rho}_f(t_f)$[15]. A history is defined as

$$H\left(\hat{\rho}_i, \hat{\Pi}_{i_1}^{(1)}, \hat{\Pi}_{i_2}^{(2)}, \ldots, \hat{\Pi}_{i_k}^{(k)}, \hat{\rho}_f\right) \equiv \hat{\rho}_i \rightarrow \hat{\Pi}_{i_1}^{(1)}(t_1) \rightarrow \hat{\Pi}_{i_2}^{(2)}(t_2) \rightarrow \cdots \rightarrow \hat{\Pi}_{i_k}^{(k)}(t_k) \rightarrow \hat{\rho}_f(t_f). \tag{78}$$

The history means that a quantum system is initialised in a state $\hat{\rho}_i$ at time t_0; at time t_1, the system has a property corresponding to $\hat{\Pi}_{i_1}^{(1)}(t_1)$; at time t_2, the system has a property corresponding to $\hat{\Pi}_{i_2}^{(2)}(t_2)$; and so on, until, at the final time t_f, the system has a property corresponding to $\hat{\rho}_f$. Given an initial state $\hat{\rho}_i$ and a final observable $\hat{\rho}_f(t_f)$, one can write down several potential histories. Two histories, $H(\hat{\rho}_i, \hat{\Pi}_{i_1}^{(1)}, \hat{\Pi}_{i_2}^{(2)}, \ldots, \hat{\Pi}_{i_k}^{(k)}, \hat{\rho}_f)$ and $H^\star(\hat{\rho}_i, \hat{\Pi}_{i_1^\star}^{(1)}, \hat{\Pi}_{i_2^\star}^{(2)}, \ldots, \hat{\Pi}_{i_k^\star}^{(k)}, \hat{\rho}_f)$, are said to be consistent with respect to one another if, and only if, their (normalised) Hilbert–Schmidt inner product vanishes:

$$\tilde{Q}(H, H^\star) \equiv \frac{1}{\text{Tr}\left(\hat{\rho}_f(t_f)\hat{\rho}_i\right)} \text{Tr}\left(\hat{\rho}_f(t_f) \hat{\Pi}_{i_k}^{(k)}(t_k) \cdots \hat{\Pi}_{i_1}^{(1)}(t_1) \hat{\rho}_i \hat{\Pi}_{i_1^\star}^{(1)}(t_1) \cdots \hat{\Pi}_{i_k^\star}^{(k)}(t_k) \hat{\rho}_f(t_f)\right) = 0. \tag{79}$$

$\tilde{Q}(H, H^\star)$ is a KD quasi-probability representing the strength of the quantum interference between two paths (histories) in Hilbert space.

A *family* $f = \{H_i\}$ of histories consists of all the histories

$$f \equiv \left\{\hat{\rho}_i \rightarrow \mathcal{C}\left(\mathcal{A}^{(1)}\right) \rightarrow \mathcal{C}\left(\mathcal{A}^{(2)}\right) \rightarrow \cdots \rightarrow \mathcal{C}\left(\mathcal{A}^{(k)}\right) \rightarrow \hat{\rho}_f\right\}, \tag{80}$$

where $\mathcal{C}\left(\mathcal{A}^{(l)}\right)$ represents any projector in $\mathcal{A}^{(l)}$. A family $f = \{H_i\}$ of histories is consistent if, and only if,

$$Q\left(H_i, H_j\right) = 0 \text{ and } H_i, H_j \in f, \ \forall i \neq j. \tag{81}$$

If f is consistent, then $\tilde{Q}(H_i, H_i)$ is a classical conditional joint probability distribution for all $H_i \in f$. Within the context of a consistent family, one can ask if an individual history has happened. Given consistency, one can regard $\tilde{Q}(H_i, H_i)$ as the probability that H_i happened.

Often, one is interested only in whether a single history happened. In this case, one need only consider the minimal family of which that history can be a member. The minimal family of $H(\hat{\rho}_i, \hat{\Pi}_{i_1}^{(1)}, \hat{\Pi}_{i_2}^{(2)}, \ldots, \hat{\Pi}_{i_k}^{(k)}, \hat{\rho}_f)$ is
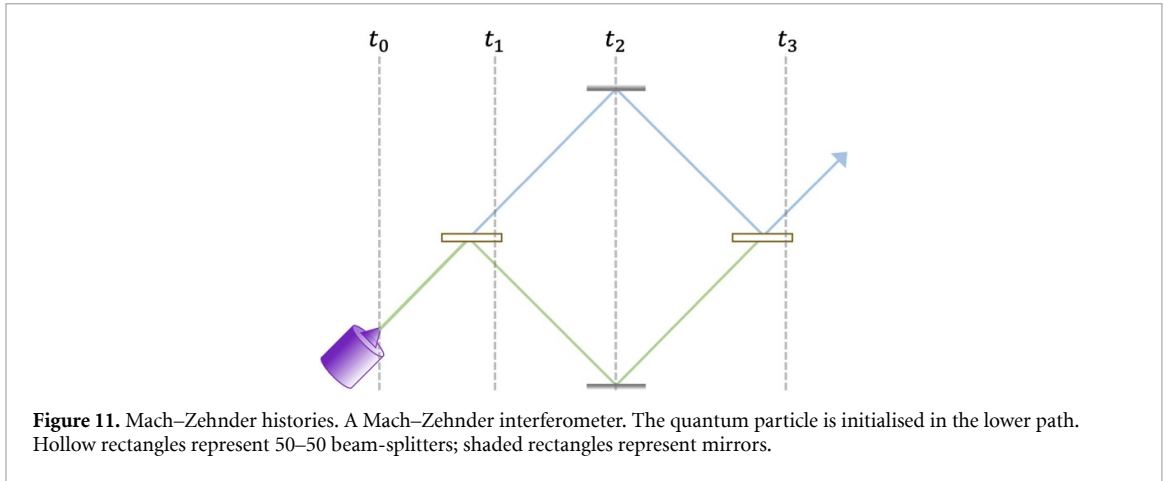
$$f_{\min}(H) \equiv \left\{\hat{\rho}_i \rightarrow \mathcal{C}\left(\left\{\hat{\Pi}_{i_1}^{(1)}, \hat{\mathbb{1}} - \hat{\Pi}_{i_1}^{(1)}\right\}\right) \rightarrow \mathcal{C}\left(\left\{\hat{\Pi}_{i_2}^{(2)}, \hat{\mathbb{1}} - \hat{\Pi}_{i_2}^{(2)}\right\}\right) \rightarrow \cdots \rightarrow \mathcal{C}\left(\left\{\hat{\Pi}_{i_k}^{(k)}, \hat{\mathbb{1}} - \hat{\Pi}_{i_k}^{(k)}\right\}\right) \rightarrow \hat{\rho}_f\right\}. \tag{82}$$

If $f_{\min}(H)$ is consistent, then the history H is also said to be consistent. In this case, one can meaningfully ask, *did H occur, or not?*

To illustrate the consistent-histories interpretation, we provide an example. Consider a Mach–Zehnder interferometer [240, 241] as shown in figure 11 . We associate the lower paths (green) with the state $|0\rangle$ and the upper paths (blue) with the state $|1\rangle$. At time t_0, a quantum particle is prepared in the state $\rho_i = |0\rangle\langle 0|$. Just before time t_1, a 50–50 beam-splitter implements the unitary operator $\hat{U}_{BS}(t_0, t_1) = e^{-i\frac{\pi}{4}\hat{\sigma}_y}$. Just before

---

[14] The theory accommodates more-general positive-operator-valued measures, however. Via Steinspring dilation, a positive-operator-valued measure defined one Hilbert space is equivalent to a projector-valued measure on a larger Hilbert space. (Non-unique dilations are problematic for ontic interpretations.).

[15] Generalisations of this setting are described in [9, 239].

Figure 11. Mach–Zehnder histories. A Mach–Zehnder interferometer. The quantum particle is initialised in the lower path. Hollow rectangles represent 50–50 beam-splitters; shaded rectangles represent mirrors.

time t_3, another 50–50 beam-splitter implements $\hat{U}_{\text{BS}}(t_2, t_3) = e^{-i\frac{\pi}{4}\hat{\sigma}_y}$. In a final event, the particle has the property $\rho_f(t_4) = |1\rangle\langle 1|$ at time t_4.

We wish to say something about the particle's position property at time t_2, between the two beam-splitters. According to the Copenhagen interpretation, we cannot say anything about this property at t_2, because no measurement was made. Let us now analyse the problem from the consistent-histories perspective. We want to ask, at t_2, if the particle had the property $\hat{\Pi}^0 \equiv |0\rangle\langle 0|$ (was on the lower path) or had the property $\hat{\Pi}^1 \equiv |1\rangle\langle 1| = \hat{\mathbb{1}} - \hat{\Pi}^0$ (was on the upper path). Define the histories

$$H_0 \equiv \hat{\rho}_i \rightarrow \hat{\Pi}^0(t_2) \rightarrow \hat{\rho}_f(t_3) \quad \text{and} \tag{83}$$

$$H_1 \equiv \hat{\rho}_i \rightarrow \hat{\Pi}^1(t_2) \rightarrow \hat{\rho}_f(t_3). \tag{84}$$

$\{H_0, H_1\}$ forms the minimal family for H_0 and for H_1; we call this family $f^{(0/1)}$. We can calculate $\tilde{Q}(H_0, H_1) = \tilde{Q}(H_1, H_0) = \frac{1}{4}$. Thus, according to the consistent-histories interpretation, one cannot ask, in a meaningful way, if the particle was in the upper or lower path at t_2.

Instead, let us define $|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$. We now ask whether, at t_2, the particle had the property $\hat{\Pi}^+ \equiv |+\rangle\langle +|$ or the property $\hat{\Pi}^- \equiv |-\rangle\langle -| = \hat{\mathbb{1}} - \hat{\Pi}^+$. Each of the histories

$$H_+ \equiv \hat{\rho}_i \rightarrow \hat{\Pi}^+(t_2) \rightarrow \hat{\rho}_f(t_3) \quad \text{and} \tag{85}$$

$$H_- \equiv \hat{\rho}_i \rightarrow \hat{\Pi}^-(t_2) \rightarrow \hat{\rho}_f(t_3) \tag{86}$$

has the minimal family $f^{(+/-)} \equiv \{H_+, H_-\}$. We can calculate $\tilde{Q}(H_+, H_-) = \tilde{Q}(H_-, H_+) = 0$. Thus, according to the consistent-histories interpretation, one can ask meaningfully if the particle had the property of a $|+\rangle$-type or a $|-\rangle$-type superposition at t_2. In fact, $\tilde{Q}(H_+, H_+) = 1$. Therefore, the consistent-histories interpretation implies that a particle initially in $\rho_i = |0\rangle\langle 0|$ and finally in $\rho_f = |1\rangle\langle 1|$ definitely had the property $\hat{\Pi}^+ = |+\rangle\langle +|$ at t_2.

8.3. Contextuality

According to a useful notion of non-classicality, an experiment is called non-classical when explaining it requires radical departures from the classical world-view [242]. Bell non-locality [243] is arguably the strongest notion of non-classicality. However, it can be provably identified only in experiments involving space-like-separated systems. A more widely applicable notion is contextuality, first developed by Kochen and Specker [244–246]. Spekkens generalised the notion later [247, 248]. Below, we review Kochen–Specker contextuality and generalised contextuality and their connections with KD distributions.

8.3.1. Kochen–Specker contextuality

In classical physics, one can ascribe exact physical properties to objects. For example, even if all measuring rods' imperfections lead to uncertainty in the measurement of a room's length, one assumes that the room has an exact length. However, a property of a quantum object, as registered by a measurement, may depend on the measurement's context—on which other properties were measured simultaneously. This dependence is *contextuality*.

The initial demonstration of Kochen–Specker contextuality involved sharp measurements of rank-1 projectors. Consider a set of rank-1 projector observables: $\mathcal{G} \equiv \{\hat{A}^{(1)}, \hat{A}^{(2)}, \ldots, \hat{A}^{(n)}\}$. Denote the ith

eigenvalue of projector j by $a_{i_j}^{(j)} \in \{0,1\}$. A Kochen–Specker context c_{KS} is a subset of compatible observables in $\mathcal{G}$, i.e. $c_{\mathrm{KS}} \subseteq \mathcal{G}$. For any given context c_{KS}, there is a probability distribution over the eigenvalues $a_{i_j}^{(j)}$ of each observable $\hat{A}^{(j)} \in c_{\mathrm{KS}}$. As all observables in c_{KS} are compatible, one can construct a joint probability consistent with the observables' individual probabilities: $P_{c_{\mathrm{KS}}}(\{a_{i_j}^{(j)} \mid \hat{A}^{(j)} \in c_{\mathrm{KS}}\})$. An experiment is Kochen–Specker non-contextual if there exists a global probability assignment over the products of the outcomes of all the observables in $\mathcal{G}$, such that it is compatible with the outcome probabilities $P_{c_{\mathrm{KS}}}(\{a_{i_j}^{(j)} \mid \hat{A}^{(j)} \in c_{\mathrm{KS}}\})$ of any considered context c_{KS}. If not, the experiment is Kochen–Specker contextual.

One might wish to ascribe to a quantum system exact physical properties, corresponding to the observables, as in classical physics. Which necessary properties of $P_{c_{\mathrm{KS}}}(\{a_{i_j}^{(j)} \mid \hat{A}^{(j)} \in c_{\mathrm{KS}}\})$ enable such an ascription? Consider contexts c_{KS} and c'_{KS} associated to probability distributions $P_{c_{\mathrm{KS}}}$ and $P_{c'_{\mathrm{KS}}}$. Consider marginalising $P_{c_{\mathrm{KS}}}$ over the outcomes of the observables absent from c'_{KS}. We denote the marginal by $\tilde{P}_{c_{\mathrm{KS}}}$, and we define $\tilde{P}_{c'_{\mathrm{KS}}}$ analogously. In terms of these distributions, we can state a condition necessary for the existence of a global probability assignment over the observables in $\mathcal{G}$ [249]:

$$\tilde{P}_{c_{\mathrm{KS}}}\left(\left\{a_{i_j}^{(j)} \mid \hat{A}^{(j)} \in c_{\mathrm{KS}} \cap c'_{\mathrm{KS}}\right\}\right) = \tilde{P}_{c'_{\mathrm{KS}}}\left(\left\{a_{i_j}^{(j)} \mid \hat{A}^{(j)} \in c_{\mathrm{KS}} \cap c'_{\mathrm{KS}}\right\}\right), \quad \forall c_{\mathrm{KS}}, c'_{\mathrm{KS}}. \tag{87}$$

We now specify the forms assumed by the joint probabilities $P_{c_{\mathrm{KS}}}(\{a_{i_j}^{(j)} \mid \hat{A}^{(j)} \in c_{\mathrm{KS}}\})$ in any Kochen–Specker–non-contextual model. Consider measuring any observable. The outcome follows deterministically from a hidden variable, regardless of the context, in any non-contextual Kochen–Specker theory. This hidden variable is called an *ontic state* and denoted by $\lambda \in \Lambda$. Λ is a discrete or continuous set of ontic states. Now, consider measuring all the observables in a context c_{KS}. We obtain the outcomes $\{a_{i_j}^{(j)} \mid \hat{A}^{(j)} \in c_{\mathrm{KS}}\}$, in any non-contextual Kochen–Specker theory, with a probability

$$P_{c_{\mathrm{KS}}}\left(\left\{a_{i_j}^{(j)} \mid \hat{A}^{(j)} \in c_{\mathrm{KS}}\right\}\right) = \sum_{\lambda} P(\lambda) P\left(\left\{a_{i_j}^{(j)} \mid \hat{A}^{(j)} \in c_{\mathrm{KS}}\right\} \mid \lambda\right). \tag{88}$$

The conditional probability factorises:

$$P\left(\left\{a_{i_j}^{(j)} \mid \hat{A}^{(j)} \in c_{\mathrm{KS}}\right\} \mid \lambda\right) = \prod_{a_{i_j}^{(j)} \,:\, \hat{A}^{(j)} \in c_{\mathrm{KS}}} P\left(a_{i_j}^{(j)} \mid \lambda\right) \in \{0,1\}. \tag{89}$$

The reason $P_{c_{\mathrm{KS}}}(\{a_{i_j}^{(j)} \mid \hat{A}^{(j)} \in c_{\mathrm{KS}}\})$ can be expressed as a product is that $P(a_{i_j}^{(j)} \mid \lambda) \in \{0,1\}$, because Kochen–Specker theory concerns deterministic hidden-variable models.

Kochen and Specker introduced example contexts in which a quantum system has properties represented by projectors. If the system is in one of certain quantum states, no definite (0 or 1) value can be ascribed to those properties [245]. Hence quantum theory is contextual in the Kochen–Specker sense. Other researchers simplified the initial proofs [250–255]. Whilst quantum theory is Kochen–Specker contextual, certain quantum experiments can be explained by Kochen–Specker non-contextual ontic models. In such experiments, one can ascribe definite properties to projective observables of a system, irrespectively of how these properties may be measured. With respect to the theory of Kochen and Specker, these experiments are deemed classical.

KD distributions offer promise for elucidating Kochen–Specker contextuality and the logical paradoxes stemming from it. Many aspects of Kochen–Specker contextuality connect to weak values (see section 4) and thus indirectly to KD distributions [256–260]. These connections invite further exploration. Here, we review the Kochen–Specker inequality proved by Klyachko, Can, Binicioğlu and Shumovsky (KCBS) [246, 255]. We highlight its close connection with KD distributions.
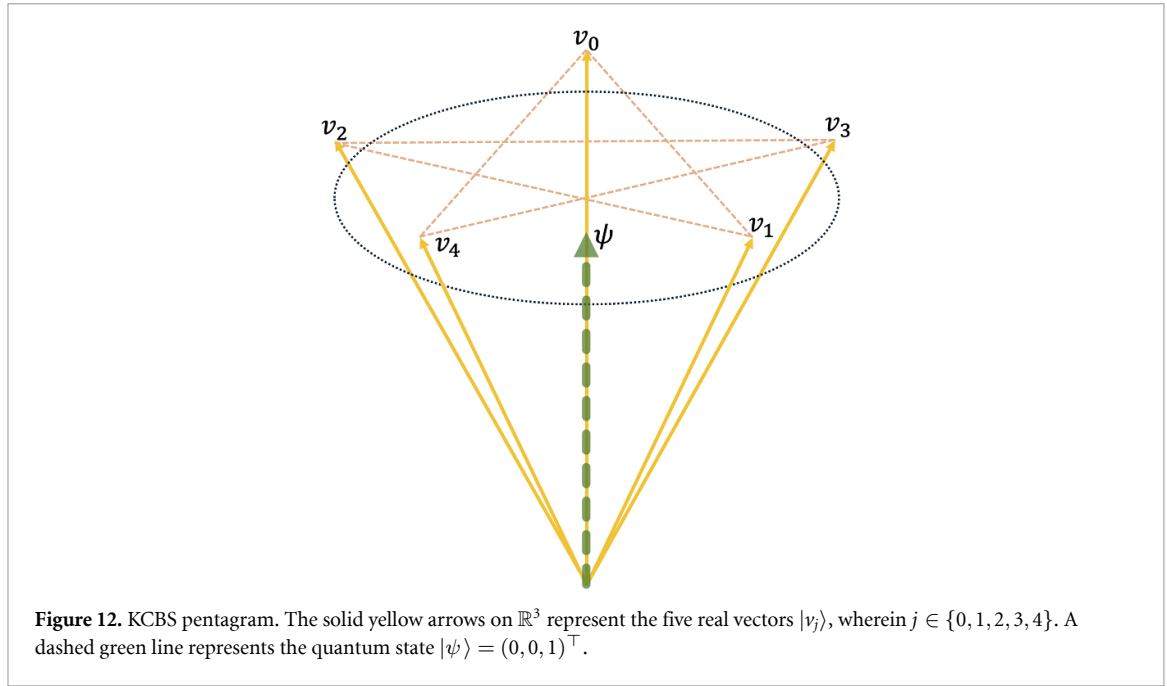
The setup includes a Hilbert space of dimensionality $d = 3$, the least dimensionality in which one can prove Kochen–Specker contextuality [246]. Consider 5 real qutrit vectors $|v_j\rangle$, wherein $j \in \{0,1,2,3,4\}$. Relative to a computational basis, $|v_j\rangle$ is represented by a list of three real numbers:

$$|v_j\rangle = \left(\sin(\theta)\sin(\phi_j),\, \sin(\theta)\cos(\phi_j),\, \cos(\theta)\right)^{\top}, \tag{90}$$

where $\cos(\theta) = \frac{1}{\sqrt[4]{5}}$ and $\phi_j = (2j+1 \mod 5) \times \frac{2\pi}{5}$. We can regard equation (90) as specifying the x-, y-, and z-coordinates of a point in three-dimensional space. In this way, we depict $|v_{0,\ldots,4}\rangle$ with solid yellow lines in the KCBS pentagram of figure 12.

From the vectors, we form five observables:

$$\hat{A}^{(j)} \equiv 2|v_j\rangle\langle v_j| - \hat{\mathbb{1}}. \tag{91}$$

Figure 12. KCBS pentagram. The solid yellow arrows on $\mathbb{R}^3$ represent the five real vectors $|v_j\rangle$, wherein $j \in \{0,1,2,3,4\}$. A dashed green line represents the quantum state $|\psi\rangle = (0,0,1)^\top$.

Above, we defined the $\hat{A}^{(j)}$ as projectors, whose eigenvalues lie in $\{0,1\}$. To simplify later algebra, we have rescaled and shifted the spectra: the eigenvalues $a_{i_j}^{(j)}$ of $\hat{A}^{(j)}$ are $+1, -1$, and -1, for all j. The observables share another important property: consider any two vectors connected by a salmon dashed line in figure 12. The corresponding vectors are orthogonal: $[\hat{A}^{(j)}, \hat{A}^{(j+1 \mod 5)}] = 0$ for all j.

KCBS introduced a set of five contexts, each formed from two commuting (compatible) observables: $\{(\hat{A}^{(0)}, \hat{A}^{(1)}), (\hat{A}^{(1)}, \hat{A}^{(2)}), (\hat{A}^{(2)}, \hat{A}^{(3)}), (\hat{A}^{(3)}, \hat{A}^{(4)}), (\hat{A}^{(4)}, \hat{A}^{(0)})\}$. Using these contexts, one can define a correlation function:

$$S \equiv \langle \hat{A}^{(0)}\hat{A}^{(1)}\rangle + \langle \hat{A}^{(1)}\hat{A}^{(2)}\rangle + \langle \hat{A}^{(2)}\hat{A}^{(3)}\rangle + \langle \hat{A}^{(3)}\hat{A}^{(4)}\rangle + \langle \hat{A}^{(4)}\hat{A}^{(0)}\rangle. \tag{92}$$

KCBS bounded S, using properties of the observables' eigenvalues, as follows. Recall that each eigenvalue $a_{i_j}^{(j)} = \pm 1$. Therefore, $(a_{i_j}^{(j)})^2 = 1$, and $\left(a_{i_0}^{(0)}a_{i_1}^{(1)}\right)\left(a_{i_1}^{(1)}a_{i_2}^{(2)}\right)\left(a_{i_2}^{(2)}a_{i_3}^{(3)}\right)\left(a_{i_3}^{(3)}a_{i_4}^{(4)}\right)\left(a_{i_4}^{(4)}a_{i_0}^{(0)}\right) = 1.$ [255] Furthermore, since each $a_{i_j}^{(j)} = \pm 1$, at least one parenthesised factor equals 1. Hence, if we exchange parentheses for pluses,

$$a_{i_0}^{(0)}a_{i_1}^{(1)} + a_{i_1}^{(1)}a_{i_2}^{(2)} + a_{i_2}^{(2)}a_{i_3}^{(3)} + a_{i_3}^{(3)}a_{i_4}^{(4)} + a_{i_4}^{(4)}a_{i_0}^{(0)} \geqslant -3. \tag{93}$$

The sum saturates the bound if 4 of the terms (the greatest possible number of terms) equal -1. Let us combine this inequality with equations (89), (92), and the upper bound of 1 on each probability:

$$S = \sum_{i_0,\dots,i_4} P\left(a_{i_0}^{(0)}, a_{i_1}^{(1)}, a_{i_2}^{(2)}, a_{i_3}^{(3)}, a_{i_4}^{(4)} | \lambda\right)\left(a_{i_0}^{(0)}a_{i_1}^{(1)} + a_{i_1}^{(1)}a_{i_2}^{(2)} + a_{i_2}^{(2)}a_{i_3}^{(3)} + a_{i_3}^{(3)}a_{i_4}^{(4)} + a_{i_4}^{(4)}a_{i_0}^{(0)}\right) \geqslant -3 \tag{94}$$

in every non-contextual deterministic hidden-variable model.

One can rephrase equation (92) in terms of a KD distribution: in terms of the five-extended KD quasi-probability

$$Q_{i_0, i_1, i_2, i_3, i_4}(\hat{\rho}) \equiv \langle a_{i_4}^{(4)}|a_{i_3}^{(3)}\rangle\langle a_{i_3}^{(3)}|a_{i_2}^{(2)}\rangle\langle a_{i_2}^{(2)}|a_{i_1}^{(1)}\rangle\langle a_{i_1}^{(1)}|a_{i_0}^{(0)}\rangle\langle a_{i_0}^{(0)}|\hat{\rho}|a_{i_4}^{(4)}\rangle, \tag{95}$$

the correlator is

$$S(Q(\hat{\rho})) = \sum_{i_0,\dots,i_4} Q_{i_0,\dots,i_4}(\hat{\rho})\left(a_{i_0}^{(0)}a_{i_1}^{(1)} + a_{i_1}^{(1)}a_{i_2}^{(2)} + a_{i_2}^{(2)}a_{i_3}^{(3)} + a_{i_3}^{(3)}a_{i_4}^{(4)} + a_{i_4}^{(4)}a_{i_0}^{(0)}\right). \tag{96}$$

Restricting $Q(\hat{\rho})$ to be a probability distribution, one reproduces the non-contextuality bound equation (94):

$$\max_{Q_{i_0,\dots,i_4} \in [0,1]} S(Q(\hat{\rho})) \geqslant -3. \tag{97}$$

Thus, if the KD distribution $Q_{i_0,\ldots,i_4}(\hat{\rho})$ is positive, then $S \geqslant -3$, and a Kochen–Specker non-contextual model describes the state. A non-positive KD distribution may break the bound. For example, define $|\psi\rangle = (1,0,0)^{\top}$ and $\hat{\rho} = |\psi\rangle\langle\psi|$. The quasi-probability $Q_{i_0,\ldots,i_4}(\hat{\rho})$ assumes negative values and violates the non-contextual inequality:

$$S(Q(\hat{\rho})) = 5 - 4\sqrt{5} \approx -3.9. \tag{98}$$

To summarise, non-positive quasi-probabilities in the KD distribution $Q_{i_0,\ldots,i_4}(\hat{\rho})$ are necessary for breaking the Kochen–Specker non-contextual inequality of *Klyachko et al*

8.3.2. Generalised contextuality

Kochen–Spekker contextuality can be ascribed only to deterministic hidden-variable models of quantum experiments that involve projective measurements. Spekkens lifted these restrictions in two ways, in developing generalised contextuality [248]. First, any ontological model may have generalised contextuality—not only deterministic hidden-variable models of quantum mechanics. Second, preparations, general measurements, and transformations can be contextual—not only projective measurements. Below, we introduce generalised contextuality. We then explain that, if a particular protocol for measuring the KD distribution yields non-positive values, the protocol is contextual. Afterwards, we discuss how KD distributions witness generalised contextuality in enhanced work extraction and in breaking bounds on thermodynamic currents.

Non-contextuality, in its generalised form [247, 248], is the property that a physical theory does not introduce distinct mathematical representations of operationally indistinguishable phenomena within a class of theories known as *ontological models*. In other words, if two things cannot be distinguished, they are the same thing within the ontological model. For example, Einstein, in his 'On the Electrodynamics of Moving Bodies' [261], criticised the way in which electrodynamics (as understood at the time via the notion of ether) assigns distinct physical descriptions to a metallic ring moving towards a magnet and to a magnet moving towards a ring. After all, the two experiments are impossible to tell apart by any measurement. One can argue that the theory-selection principle being applied is non-contextuality [262].

The word *ontological* labels a theory of an underlying physical reality. The underlying physical reality may be inaccessible (hidden). An experiment consists of preparation procedures $\mathscr{P}$, transformation procedures $\mathscr{T}$, and measurement procedures $\mathscr{M}$. These generate a set $\{P(k|\mathscr{P},\mathscr{T},\mathscr{M})\}_{\mathscr{P},\mathscr{T},\mathscr{M}}$ of outcomes statistics. $p(k|\mathscr{P},\mathscr{T},\mathscr{M})$ denotes the probability of observing outcome k, given that we prepared the system according to $\mathscr{P}$, transformed it according to $\mathscr{T}$ and measured it according to $\mathscr{M}$. An ontological model associates to operational procedures (preparations, transformations and measurements) probabilistic representations over a measurable state space Λ [247], the set of *physical states*. We model a preparation procedure $\mathscr{P}$ by sampling $\lambda \in \Lambda$ according to a probability distribution $P_{\mathscr{P}}(\lambda)$. Each transformation $\mathscr{T}$ is modelled by a matrix; each element $P_{\mathscr{T}}(\lambda'|\lambda)$ equals the probability of jumping from λ to λ'. Each measurement procedure $\mathscr{M}$ is modelled by another matrix; each element $P_{\mathscr{M}}(k|\lambda')$ equals the probability of obtaining outcome k, given the ontic state λ'.

The maps $\mathscr{P} \mapsto P_{\mathscr{P}}$, $\mathscr{T} \mapsto P_{\mathscr{T}}$ and $\mathscr{M} \mapsto P_{\mathscr{M}}$ must be convex-linear. This property ensures a reasonable probabilistic interpretation. For example, imagine tossing a fair coin to decide whether we follow the preparation procedure $\mathscr{P}_1$ or $\mathscr{P}_2$. The ontological model must assign the probability $[P_{\mathscr{P}_1}(\lambda) + P_{\mathscr{P}_2}(\lambda)]/2$ to the physical state λ. From the propagation of probabilities, an ontological model predicts that outcome k will occur with a probability

$$P(k|\mathscr{P},\mathscr{T},\mathscr{M}) = \int_{\Lambda} \mathrm{d}\lambda \int_{\Lambda} \mathrm{d}\lambda'\, P_{\mathscr{M}}(k|\lambda')\, P_{\mathscr{T}}(\lambda'|\lambda)\, P_{\mathscr{P}}(\lambda). \tag{99}$$

See [263] for an introduction to ontological models.

Ontological models, unless otherwise restricted, can reproduce nearly anything. For example, consider a quantum experiment. In the quantum formalism, $\mathscr{P}$ is represented by a pure state $|\psi\rangle \in \mathbb{C}^d$, $\mathscr{T}$ by a $d \times d$ unitary $\hat{U}$ and $\mathscr{M}$ by a projective measurement $\mathcal{M} = \{|k\rangle\langle k|\}_{k=1}^d$. By the Born rule, $P(k|\mathscr{P},\mathscr{T},\mathscr{M}) = |\langle k|\hat{U}|\psi\rangle|^2$. The right-hand side of equation (99) can reproduce this function if Λ consists of normalised pure states. In this model, whose ontic states are normalised pure states, the following assignments are made: $P_{\psi}(\lambda) = \delta(\lambda - \psi)$, $P_{\mathscr{T}}(\lambda'|\lambda) = \delta(\lambda' - \hat{U}\psi)$ and $P_{\mathscr{M}}(k|\lambda') = |\langle k|\lambda'\rangle|^2$ [264]. On can represent, within the ontological model, a general preparation procedure $\mathscr{P}$ with a probability measure on Λ:

$$P_{\mathscr{P}}(\lambda) = \sum_n P_n \delta(\lambda - \psi_n) = \sum_n P_n P_{\psi_n}(\lambda), \quad \text{such that} \quad \sum_n P_n = 1. \tag{100}$$

Hence a general ontological model can model any quantum experiment. Ontological models need not be deterministic. For example, in the above model, λ determines the measurement outcome probabilistically.

Non-contextuality can restrict ontological models to satisfy a notion of classicality. Imagine preparing two copies of a system according to two different procedures, $\mathscr{P}$ and $\mathscr{P}'$. Suppose that we obtain the same measurement statistics, no matter which measurements $\mathscr{M}$ we conduct: $P(k|\mathscr{P},\mathscr{M}) = P(k|\mathscr{P}',\mathscr{M})$ for all k and $\mathscr{M}$. This condition, known as an *operational equivalence*, is denoted by $\mathscr{P} \sim \mathscr{P}'$ [248]. Non-contextuality requires that the ontological model assign the same distribution over λ to both preparations: $P_{\mathscr{P}}(\lambda) = P_{\mathscr{P}'}(\lambda)$ for all λ. For example, consider forming a non-contextual description of the preparation of a density operator $\hat{\rho}$. We must represent $\hat{\rho}$ by the same $P_{\hat{\rho}}(\lambda)$, independently of the pure-state mixture used to prepare $\hat{\rho}$. The previous example of an ontological model for quantum mechanics is contextual (is not non-contextual). Indeed, $P_{\mathscr{P}}$ can differ from $P_{\mathscr{P}'}$ (see equation (100)), while

$$\hat{\rho}_{\mathscr{P}} = \sum_n P_n |\psi_n\rangle\langle\psi_n| = \hat{\rho}_{\mathscr{P}'} = \sum_n P_n' |\psi_n'\rangle\langle\psi_n'|. \tag{101}$$

In other words, in contextual models, different procedures $\mathscr{P}$ can yield different $P_{\mathscr{P}}$, while nevertheless yielding the same density matrix (see equation (100)) and hence the same experimental outcomes. In non-contextual models $\mathscr{P} \sim \mathscr{P}'$ implies $P_{\mathscr{P}}(\lambda) = P_{\mathscr{P}'}(\lambda)$. Similarly, consider two transformations that realise the same quantum channel $\mathcal{C}$. Non-contextuality requires that the transformations be assigned the same $P_{\mathcal{C}}(\lambda'|\lambda)$. A similar rule governs two measurements associated to the same positive-operator-valued measure $\mathcal{M}$. The measurements are assigned the same $P_{\mathcal{M}}(k|\lambda)$ in any non-contextual model. In short, generalised non-contextuality is the property that operational equivalence implies ontological identity.

Generalised non-contextuality relates to Kochen–Specker contextuality (outlined above) as follows. Consider a sharp measurement represented, in quantum theory, by the projector-valued measures $\mathcal{M} = \{\hat{\Pi}_k\}$. Sharp measurements are commonly assumed to obey outcome determinism in classical theories [248]. That is, the ontic state λ is assumed to dictate a measurement's outcome deterministically: $P_{\mathcal{M}}(k|\lambda) \in \{0,1\}$. Consider sharp measurements whose projector-valued measures, $\mathcal{M}_1$ and $\mathcal{M}_2$, share a projector $\hat{\Pi}_{k^\star}: \hat{\Pi}_{k^\star} \in \mathcal{M}_1 \cap \mathcal{M}_2$. By the above definitions of generalised measurement non-contextuality,

$$P_{\mathcal{M}_1}(k^\star|\lambda) = P_{\mathcal{M}_2}(k^\star|\lambda). \tag{102}$$

As $k^\star$ labels a projector $\hat{\Pi}_{k^\star}$, outcome determinism ensures that [153]

$$P_{\mathcal{M}_1}(k^\star|\lambda) = P_{\mathcal{M}_2}(k^\star|\lambda) \in \{0,1\}. \tag{103}$$

The first equality implies that the probability of observing $k^\star$ does not depend on the context within which the projector is measured (does not depend on whether $\mathcal{M}_1$ or $\mathcal{M}_2$ is measured). The second equality states that the $k^\star$ outcome's probability equals 0 or 1, depending on the ontic state λ. The present problem satisfies the definition of Kochen–Specker non-contextuality (equation (89), etc). To summarise, under the assumption of outcome determinism and sharp measurements, generalised measurement non-contextuality implies Kochen–Specker non-contextuality.

From an experimental viewpoint, operational equivalence need not be checked exactly. In fact, they cannot be. Fortunately, contextuality proofs can accommodate experimental imperfections, e.g. [265–267]. Experimental contextuality tests must check operational equivalences for a tomographically complete set of preparations and measurements. To achieve this complete testing one can rely on quantum mechanics to tell us what these tomographically complete sets are. Alternatively, we can perform theory-agnostic generalised-probabilistic-theory tomography [268, 269]. For other questions related to generalised contextuality and its experimental testing, see [270].

If an experiment $\{P(k|\mathscr{P},\mathscr{T},\mathscr{M})\}_{\mathscr{P},\mathscr{T},\mathscr{M}}$ is *contextual*, no non-contextual ontological model satisfies equation (99). Contextuality has been identified in (specific aspects of) experiments that involve state discrimination [265, 271], approximate cloning [266], uncertainty relations [272], interference [271], randomness certification [273] and more. However, we focus on the link between contextuality and quasi-probabilities—specifically, the KD distribution.

A positive quasi-probability representation of quantum mechanics (or of a subtheory of quantum mechanics) assigns to each density operator $\hat{\rho}$, quantum channel $\mathcal{C}$ and positive-operator-valued measure $\mathcal{M} = \{\hat{M}_k\}$ probability distributions $P_{\hat{\rho}}(\lambda)$, $P_{\mathcal{C}}(\lambda'|\lambda)$ and $P_{\mathcal{M}}(k|\lambda)$ over a measurable space Λ [37]. Density operators, channels and positive-operator-valued measures can be understood as equivalence classes of preparation procedures, transformations and measurements. (The *equivalence* refers to the above sense of

operational equivalence.) Hence, two conditions are equivalent: the existence of a non-contextual ontological model for an experiment and the existence of a positive quasi-probability representation of the experiment [11]. In other words, contextuality is equivalent to the negativity or non-reality of *every* quasi-probability representation of the experiment. A particular quasi-probability's negativity does not suffice for proving contextuality, generally. (However, the condition can suffice in some subtheories of quantum mechanics [274].) Nevertheless, there exist surprisingly direct connections between the KD distribution and generalised contextuality.

Suppose that a KD distribution has negative real part:

$$\text{Re}\left(Q_{j,k}(\hat{\rho})\right) = \text{Re}\left(\text{Tr}\left(\hat{\Pi}_k^{(B)}\hat{\Pi}_j^{(A)}\hat{\rho}\right)\right) < 0. \tag{104}$$

The following sequential-measurement scheme (the weak-value experiment described in section 4.1) is then contextual [267, 275]: weakly measure $\{\hat{\Pi}_j^{(A)}\}$ (measuring the meter's position eigenbasis), then strongly measure $\{\hat{\Pi}_k^{(B)}\}$. No non-contextual ontological model can explain the set $\{P(k|\mathscr{P},\mathscr{T},\mathscr{M})\}_{\mathscr{P},\mathscr{T},\mathscr{M}}$ of outcome statistics (including the statistics that support the operational equivalences) [275]. Therefore, KD negativity implies the negativity of every quasi-probability representation of this experiment. Similarly, suppose that $\text{Im}\left(Q_{j,k}(\hat{\rho})\right) \neq 0$. An analogous scheme, in which the meter's momentum eigenbasis is measured, is contextual [267]. Even if a continuous meter's outcomes are coarse-grained (even if a qubit replaces the meter), the results hold [267]. In conclusion, KD non-positivity and contextuality characterise an experiment that reports an anomalous weak value [267, 275].

The KD distribution also witnesses contextuality in the context of work extraction (and injection), in the linear-response regime discussed in section 6 [170]. We have already mentioned that the work scales as $O(g)$ only when a the imaginary part of an average with respect to a KD distribution is non-zero. Otherwise, the work scales as $O(g^2)$. More precisely,

$$\langle W \rangle = \frac{2g\tau}{\hbar}\text{Im}\left(\text{Tr}\left(H_0\bar{V}_I(\tau)\hat{\rho}\right)\right) + O\left(g^2\right). \tag{105}$$

This expression connects the work's $O(g)$ behaviour with a KD quantity's being non-zero: $\text{Im}\left(\text{Tr}\left(H_0\bar{V}_I(\tau)\hat{\rho}\right)\right) \neq 0$. This result says nothing directly about contextuality. However, a work-extraction transformation $\mathscr{T}$ (a unitary in quantum mechanics), in the presence of an extra condition called *stochastic reversibility*, can generate $O(g)$ behaviour only if the experiment is contextual [170]. Stochastic reversibility is the condition under which one can reverse a transformation $\mathscr{T}$, up to first order in g, via probabilistic mixture with another operation $\mathscr{T}^*$. That is, there exist transformations $\mathscr{T}^*$ and $\mathscr{T}'$ such that

$$\frac{1}{2}\mathscr{T} + \frac{1}{2}\mathscr{T}^* \sim (1-p_d)I + p_d\mathscr{T}', \tag{106}$$

where I is the trivial (identity) transformation and $p_d = O(g^2)$. Qubit systems always satisfy the above operational equivalence, according to quantum mechanics.

Complementary to work is heat. As discussed in section 6, KD distributions satisfy many desirable criteria for distributions characterising measurements of fluctuating work, heat and entropy. Moreover, we have seen that KD non-positivity witnesses contextuality in experiments that report anomalous weak values. The connection between such experiments and contextuality has led to theoretical bounds on the average heat [152] and average work [158] that can flow during a sequential-measurement protocol admitting of a non-contextual ontological model. These bounds can be broken only if the sequential-measurement process is contextual. Like heat and work, entropy production is a thermodynamic quantity that fluctuates from trial to trial. A notion of stochastic entropy production can be defined via KD distributions [160]. A non-real stochastic entropy production signals the contextuality of a thermodynamic process that involves sequential measurements [160].

9. Mathematical structure of the KD-positive states and properties of KD non-positivity

If $Q(\hat{\rho})$ has neither negative nor non-real components, it is a classical probability distribution. A KD distribution's ability to assume negative and non-real values allows the distribution to describe quantum experiments that cannot be described and analysed with classical probability theories. Recall that we refer to negative or non-real values simply as *non-positive*, and that we call $Q(\hat{\rho})$ *positive* if all its entries are positive or zero. Consider a KD distribution defined with respect to two operators's eigenbases. For the KD

distribution to contain non-positive values, a necessary but insufficient condition is that the operators do not commute [32]. Thus, KD non-positivity can be seen as a property related to, but stricter than, non-commutation.

As we saw in previous sections, experiments described by non-positive KD distributions can often generate data that cannot be generated via sampling from classical joint probability distributions. It is, therefore, important to understand when a KD distribution is non-positive. In this section, we summarise the literature on this topic.

9.1. KD positivity

Consider a d-dimensional Hilbert space $\mathcal{H}$, as well as k projective positive-operator-valued measures $\mathcal{M}^{(l)} \equiv \{\hat{M}_{i_l}^{(l)}\}$, where $l = 1, 2, \ldots, k$. We call a state $\hat{\rho}$ KD-positive if, for all indices, $Q_{i_1,\ldots,i_k}(\hat{\rho}) \geqslant 0$. In this case, $Q(\hat{\rho})$ defines a probability distribution over the set of indices. KD positivity of $\hat{\rho}$ depends on the choice of positive-operator-valued measures $\mathcal{M}^{(l)}$. In what follows, we will not indicate this dependence, the choice being clear from the context.

We denote the set of all KD-positive states by $\mathcal{E}_{\mathrm{KD}+}$. $\mathcal{E}_{\mathrm{KD}+}$ is a convex, bounded, closed set. We denote by $\mathcal{E}_{\mathrm{KD}+}^{\mathrm{ext}}$ the set of extreme points[16] of $\mathcal{E}_{\mathrm{KD}+}$. Further, we denote the set of pure KD-positive states by $\mathcal{E}_{\mathrm{KD}+}^{\mathrm{pure}}$. We denote by $\mathrm{conv}(\mathcal{E})$ the set of all convex combinations of the elements of a set $\mathcal{E}$. A convex combination is defined as a weighted finite sum of elements $\hat{\rho}_j \in \mathcal{E}$, of the form $\sum_j \alpha_j \hat{\rho}_i$, such that $\sum_j \alpha_j = 1$ and $\alpha_j \geqslant 0$. One has that $\mathcal{E}_{\mathrm{KD}+}^{\mathrm{pure}} \subseteq \mathcal{E}_{\mathrm{KD}+}^{\mathrm{ext}}$, and thus

$$\mathrm{conv}\left(\mathcal{E}_{\mathrm{KD}+}^{\mathrm{pure}}\right) \subseteq \mathrm{conv}\left(\mathcal{E}_{\mathrm{KD}+}^{\mathrm{ext}}\right) = \mathcal{E}_{\mathrm{KD}+}. \tag{107}$$

The last equality follows from the Krein–Milman theorem [276].

The following lemma links positivity of the KD distribution of all states to commutativity of the positive-operator-valued measures $\mathcal{M}^{(l)} = \{\hat{M}_{i_l}^{(l)}\}$ used in the distribution's construction.

Lemma 9.1. *The two following statements are equivalent:*

- *For all quantum states $\hat{\rho}$, the KD distribution is positive: $Q(\hat{\rho}) \geqslant 0$.*
- *All positive-operator-valued-measure operators pairwise commute:* $\left[\hat{M}_{i_l}^{(l)}, \hat{M}_{i_{l'}}^{(l')}\right] = 0$ *for all l, l' and all $i_l, i_{l'}$.*

The lemma prevents any state $\hat{\rho}$ from having negative or non-real KD quasiprobabilities when all the positive operators $\hat{M}_{i_l}^{(l)}$ commute pairwise. By contraposition, the lemma also guarantees the existence of a quantum state $\hat{\rho}$ for which $Q(\hat{\rho})$ is not a probability distribution, when there exists at least one pair (l, l') and one pair $(i_l, i_{l'})$ such that $\hat{M}_{i_l}^{(l)}$ and $\hat{M}_{i_{l'}}^{(l')}$ do not commute. So non-commutativity of measurement operators, a typical quantum feature linked to incompatibility [277], is a prerequisite for the existence of at least one state $\hat{\rho}$ that is not KD-positive. Further links between incompatibility and non-commutativity of observables, uncertainty of states with respect to observables, and non-positivity of the KD distribution are elaborated upon in section 9.5 (see [32–34]).

Proof. To begin, we prove that the lemma's first point implies the second point. For ease of notation, and without loss of generality, we focus on $l = 1$ and $l' = 2$. Since $Q(\hat{\rho}) \geqslant 0$ for all quantum states $\hat{\rho}$,

$$\forall (i_1, i_2), \, \mathrm{Tr}\left(\hat{M}_{i_2}^{(2)} \hat{M}_{i_1}^{(1)} \hat{\rho}\right) \geqslant 0. \tag{108}$$

This statement governs all pure states, in particular. Therefore, $\hat{M}_{i_2}^{(2)} \hat{M}_{i_1}^{(1)} \geqslant 0$ for all (i_1, i_2). Consequently, $\hat{M}_{i_2}^{(2)} \hat{M}_{i_1}^{(1)}$ is self-adjoint. Thus,

$$\hat{M}_{i_1}^{(1)} \hat{M}_{i_2}^{(2)} = \hat{M}_{i_2}^{(2)} \hat{M}_{i_1}^{(1)}, \tag{109}$$

concluding the proof.

Now, we prove that the lemma's second point implies the first. Recall that, if $\hat{X}$ and $\hat{Y}$ are non-negative operators (if $\hat{X} \geqslant 0$ and $\hat{Y} \geqslant 0$) and if they commute, then $\hat{X}\hat{Y} \geqslant 0$. Therefore, under the hypothesis of the lemma's second point, for all $(i_1, \ldots, i_k)$, $\hat{M}_{i_k}^{(k)} \cdots \hat{M}_{i_1}^{(1)} \geqslant 0$.

$\square$

---

[16] A convex set's extreme points are the points that do not lie on any open line segment that connects two points in the set.

9.2. Characterising the set of KD-positive states

A precise, simple characterisation of the KD-positive states is not easy to come by. Most efforts have focused on the standard KD distribution (equation (14)), for which $k = 2$. We now discuss these efforts in some detail. We further suppose that the measurement operators $\hat{M}_{i_l}^{(l)}$, with $l = 1, 2$, form complete sets $\{\hat{\Pi}_{i_l}^{(l)}\}$ of one-dimensional orthogonal projectors in a d-dimensional Hilbert space. In other words, there exist two orthonormal bases, $\{|a_i\rangle\}$ and $\{|b_j\rangle\}$, such that $\hat{\Pi}_i^{(1)} = |a_i\rangle\langle a_i|$ and $\hat{\Pi}_j^{(2)} = |b_j\rangle\langle b_j|$. For simplicity, we set $i \equiv i_1$ and $j \equiv i_2$. We introduce

$$\mathcal{A} = \{|a_i\rangle\langle a_i|\} \quad \text{and} \quad \mathcal{B} = \{|b_j\rangle\langle b_j|\}. \tag{110}$$

The union of the projectors onto the basis vectors forms a subset of the pure KD-positive states, which form a subset of the extremal KD-positive states:

$$\mathcal{A} \cup \mathcal{B} \subseteq \mathcal{E}_{\mathrm{KD}+}^{\mathrm{pure}} \subseteq \mathcal{E}_{\mathrm{KD}+}^{\mathrm{ext}}. \tag{111}$$

Indeed, one can straightforwardly check that all the basis states are KD-positive. In view of equation (111), in the simplest situation,

$$\mathcal{A} \cup \mathcal{B} = \mathcal{E}_{\mathrm{KD}+}^{\mathrm{pure}} = \mathcal{E}_{\mathrm{KD}+}^{\mathrm{ext}}. \tag{112}$$

In this situation, $\mathcal{E}_{\mathrm{KD}+} = \mathrm{conv}(\mathcal{A} \cup \mathcal{B})$, and $\mathcal{E}_{\mathrm{KD}+}$ forms a polytope.

We define a unitary transition matrix $\hat{U}$ whose entries are $\hat{U}_{i,j} \equiv \langle a_i | b_j \rangle$. The following result sums up multiple situations in which equation (112) holds [35].

Theorem 9.2. *Suppose that $m_{\mathcal{A},\mathcal{B}} \equiv \min_{i,j} |\langle a_i | b_j \rangle| > 0$. Equation (112) is true under any one of the following conditions:*

1. *$d = 2$.*
2. *d is a prime number, and the transition matrix $\hat{U}$ equals the discrete-Fourier-transform matrix.*
3. *In any dimension d, for a set of transition matrices $\hat{U}$ that has probability 1 in the set of all unitary matrices.*

In the three scenarios covered by the theorem, the only extreme KD-positive states are the pure KD-positive states and the only pure KD-positive states are the basis states belonging to $\mathcal{A}$ or $\mathcal{B}$. A more detailed statement, as well as proofs of parts (1) and (2), appear in [35]. There, part (3) of the theorem is conjectured. The proof of part (3) appears in [36].

The theorem's condition $m_{\mathcal{A},\mathcal{B}} > 0$ can be interpreted as a weak form of incompatibility for the measurements associated with $\mathcal{A}$ and $\mathcal{B}$ [34]. As mentioned above, $m_{\mathcal{A},\mathcal{B}} > 0$ also implies that the KD distribution is informationally complete: It determines $\hat{\rho}$ uniquely (see equation (15)).

In specific cases, the structure of $\mathcal{E}_{\mathrm{KD}+}$ as a convex set can be considerably more complicated than when the equalities of equation (112) hold. Figure 13 shows a schematic representation of the situation where

$$\mathcal{A} \cup \mathcal{B} \subsetneq \mathcal{E}_{\mathrm{KD}+}^{\mathrm{pure}} \subsetneq \mathcal{E}_{\mathrm{KD}+}^{\mathrm{ext}}. \tag{113}$$
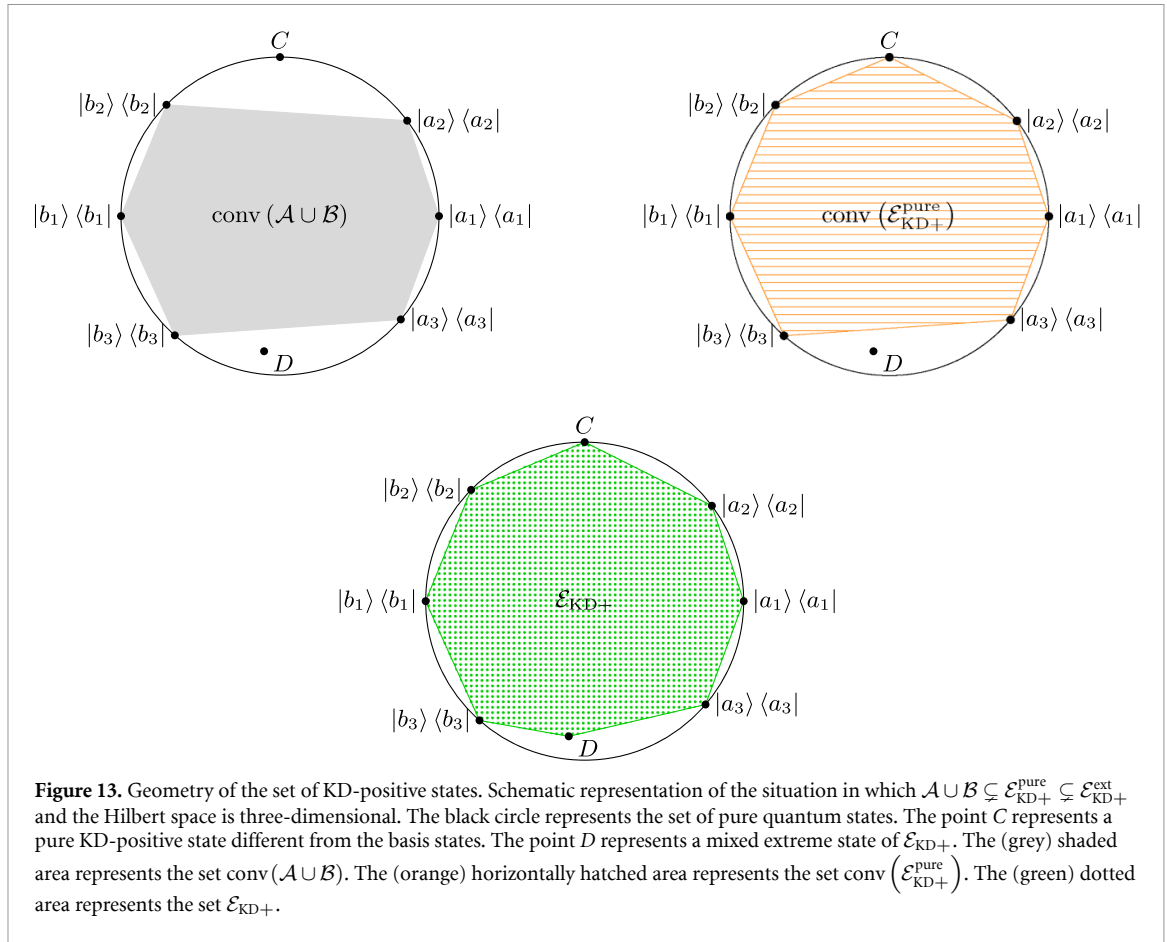
In this case, some pure KD-positive states are not basis states, and some extreme KD-positive states are mixed. The latter type of states are KD-positive, yet cannot be expressed as convex combinations of KD-positive pure states. Such states have been used to investigate fundamental aspects of generalised contextuality (see section 8.3.2). Certain experiments that measure these states' KD distributions can be described by non-contextual models, yet the experiments can verify contextuality [278]. The situation described in equation (113) arises in dimension $d = 3$, e.g. for spin-1 systems: One can choose for $\{|a_i\rangle\}$ the eigenbasis of the angular-momentum operator $\hat{J}_z$ in the z-direction and, for $\{|b_j\rangle\}$, the eigenbasis of the angular-momentum operator $\hat{J}_{\hat{n}}$ for an appropriate rotation axis $\hat{n}$ [35]. More information about $\mathcal{E}_{\mathrm{KD}+}^{\mathrm{pure}}$, when the transition matrix is a Hadamard matrix ($|\langle a_i | b_j \rangle|^2 = \frac{1}{d} = m_{\mathcal{A},\mathcal{B}}$ for all i,j), appears in [33, 34, 279].

9.3. Measure of KD non-positivity

If, and only if, a KD distribution is a classical joint probability distribution, $\sum_{i_1,\ldots,i_k} |Q_{i_1,\ldots,i_k}(\hat{\rho})| = 1$. Thus, a popular measure of a KD distribution's total non-positivity is

$$\mathcal{N}(Q_{i_1,\ldots,i_k}(\hat{\rho})) = \sum_{i_1,\ldots,i_k} |Q_{i_1,\ldots,i_k}(\hat{\rho})| \geqslant 1, \tag{114}$$

as defined in [40]. We now list a few properties of this non-positivity measure. Details appear in [32].

Figure 13. Geometry of the set of KD-positive states. Schematic representation of the situation in which $\mathcal{A} \cup \mathcal{B} \subsetneq \mathcal{E}_{\text{KD+}}^{\text{pure}} \subsetneq \mathcal{E}_{\text{KD+}}^{\text{ext}}$ and the Hilbert space is three-dimensional. The black circle represents the set of pure quantum states. The point C represents a pure KD-positive state different from the basis states. The point D represents a mixed extreme state of $\mathcal{E}_{\text{KD+}}$. The (grey) shaded area represents the set $\text{conv}(\mathcal{A} \cup \mathcal{B})$. The (orange) horizontally hatched area represents the set $\text{conv}\left(\mathcal{E}_{\text{KD+}}^{\text{pure}}\right)$. The (green) dotted area represents the set $\mathcal{E}_{\text{KD+}}$.

1. *Positivity:* a KD distribution is a classical probability distribution if, and only if, $\mathcal{N}(Q_{i_1,\ldots,i_k}(\hat{\rho})) = 1$.
2. *Stricter than non-commutation:* if $\mathcal{N}(Q_{i_1,\ldots,i_k}(\hat{\rho})) > 1$, then $[\hat{A}, \hat{B}] \neq 0$, $[\hat{A}, \hat{\rho}] \neq 0$ and $[\hat{B}, \hat{\rho}] \neq 0$. However, even if $[\hat{A}, \hat{B}] \neq 0$, $[\hat{A}, \hat{\rho}] \neq 0$ and $[\hat{B}, \hat{\rho}] \neq 0$, the corresponding KD distribution may be positive, such that $\mathcal{N}(Q_{i_1,\ldots,i_k}(\hat{\rho})) = 1$.
3. *Convexity:* the total non-positivity is a convex function with respect to mixed states: $\mathcal{N}(Q_{i_1,\ldots,i_k}\left(\sum_t p_t \hat{\rho}_t\right)) \leqslant \sum_t p_t \mathcal{N}(Q_{i_1,\ldots,i_k}(\hat{\rho}_t))$.
4. *Extension dependence:* a KD distribution $Q_{i_1,\ldots,i_l}(\hat{\rho})$ may follow from the marginalisation of another KD distribution, $Q_{i_1,\ldots,i_k}(\hat{\rho})$, where $k > l$. In this case, $\mathcal{N}(Q_{i_1,\ldots,i_l}(\hat{\rho})) > 1$ implies that $\mathcal{N}(Q_{i_1,\ldots,i_k}(\hat{\rho})) \geqslant \mathcal{N}(Q_{i_1,\ldots,i_l}(\hat{\rho})) > 1$.
5. *Coarse-graining:* a KD distribution $Q_{i_1,\ldots,i_k}(\hat{\rho})$ may represent a state $\hat{\rho}$ in terms of measurement operators that result from coarse-graining the operators in another KD distribution, $Q'_{i_{1'},\ldots,i_{k'}}(\hat{\rho})$, where $k' > k$. In this case, $\mathcal{N}(Q_{i_1,\ldots,i_k}(\hat{\rho})) \leqslant \mathcal{N}(Q'_{i_{1'},\ldots,i_{k'}}(\hat{\rho}))$.
6. *Maximum value:* consider the KD distributions defined in terms of projective positive-operator-valued measures $\mathcal{M}^{(1)}, \ldots, \mathcal{M}^{(k)}$ (see section 2.5). For such KD distributions, the maximum value of $\mathcal{N}$ depends on both the Hilbert-space dimension d and the extendedness k:

$$\max_{\hat{\rho},\mathcal{M}^{(1)},\ldots,\mathcal{M}^{(k)}} \{\mathcal{N}(Q_{i_1,\ldots,i_k}(\hat{\rho}))\} = d^{(k-1)/2}. \tag{115}$$

To elucidate when the maximum, equation (115), attains, we review *mutually unbiased bases.* Bases $\{|a_i\rangle\}$ and $\{|b_j\rangle\}$ are *mutually unbiased* if

$$|\langle a_i|b_j\rangle| = 1/\sqrt{d}, \tag{116}$$

for all i, j. For details, see [280]. Suppose that a basis $\{|a_{i_l}^{(l)}\rangle\}$ is mutually unbiased with respect to $\{|a_{i_{l\oplus1}}^{(l\oplus1)}\rangle\}$, for all $l = 1, 2, \ldots, k$. The $\oplus$ represents addition modulo k. Suppose, further, that $\hat{\rho}$ is pure and that $\langle a_{i_1}^{(1)}|\hat{\rho}|a_{i_1}^{(1)}\rangle = \langle a_{i_k}^{(k)}|\hat{\rho}|a_{i_k}^{(k)}\rangle = 1/d$ for all i_1 and i_k. This situation achieves the maximum equation (115). Conceptually, the neighbouring vectors in equation (27) should be mutually unbiased.

9.4. KD non-positivity and coherence

Quantum superpositions play an important role in the study of non-classicality. Furthermore, superpositions accompany coherence. Therefore, we will sketch a connection between KD non-positivity and quantum coherence.

We begin with background information about coherence. Consider a basis $\{|a_i\rangle\}$. In terms of it, we can express a quantum state $\hat{\rho}$ as

$$\hat{\rho} = \sum_{i,j} \rho_{i,j} |a_i\rangle\langle a_j|, \tag{117}$$

wherein $\rho_{i,j} \equiv \langle a_i|\hat{\rho}|a_j\rangle$. $\hat{\rho}$ is incoherent with respect to the basis $\{|a_i\rangle\}$ if $\rho_{i,j} = 0$ for all $i \neq j$. If $\rho_{i,j} \neq 0$ for any $i \neq j$, then $\hat{\rho}$ is coherent with respect to $\{|a_i\rangle\}$. A popular measure of quantum coherence is based on the ℓ_1 norm. Denote by $\mathcal{I}_{\{|a_i\rangle\}}$ the set of all $\{|a_i\rangle\}$-incoherent states. The ℓ_1 coherence of $\hat{\rho}$ with respect to $\{|a_i\rangle\}$ is

$$C_{\ell_1}(\hat{\rho}; \{|a_i\rangle\}) \equiv \min_{\hat{\sigma} \in \mathcal{I}_{\{|a_i\rangle\}}} \{||\hat{\rho} - \hat{\sigma}||_{\ell_1}\} = \sum_{i \neq j} |\rho_{i,j}|. \tag{118}$$

The KD distribution $Q_{i,j}(\hat{\rho}) = \langle b_j|a_i\rangle\langle a_i|\hat{\rho}|b_j\rangle$ quantifies coherence [281–283]. Consider summing the absolute values of the quasi-probabilities imaginary components, then maximising over all bases $\{|b_j\rangle\}$:

$$C_{\mathrm{KD}}(\hat{\rho}; \{|a_i\rangle\}) \equiv \max_{\{|b_j\rangle\}} \{\mathcal{N}(\mathrm{Im}Q(\hat{\rho}))\} = \max_{\{|b_j\rangle\}} \sum_{i,j} \{|\mathrm{Im}\langle b_j|a_i\rangle\langle a_i|\hat{\rho}|b_j\rangle|\}. \tag{119}$$

$C_{\mathrm{KD}}(\hat{\rho}; \{|a_i\rangle\})$ witnesses coherence faithfully: $C_{\mathrm{KD}}(\hat{\rho}; \{|a_i\rangle\}) = 0$ if and only if $\hat{\rho}$ is incoherent in $\{|a_i\rangle\}$. Budiyono *et al* [281–283] demonstrate further relationships between $C_{\mathrm{KD}}(\hat{\rho}; \{|a_i\rangle\})$ and coherence.

Another KD distribution is more directly related to $C_{\ell_1}(\hat{\rho}; \{|a_i\rangle\})$. Consider the extended KD distribution $Q^\star_{i,j,k}(\hat{\rho}) = \langle a_j|b_k\rangle\langle b_k|a_i\rangle\langle a_i|\hat{\rho}|a_j\rangle$, wherein $\{|b_k\rangle\}$ is mutually unbiased with respect to $\{|a_i\rangle\}$: $|\langle a_i|b_k\rangle| = 1/\sqrt{d}$, for all i and j. The total non-positivity (equation (114)) minus 1 equals the ℓ_1 measure of coherence:

$$\mathcal{N}(Q^\star(\hat{\rho})) - 1 = -1 + \sum_{i,j,k} |\langle a_j|b_k\rangle\langle b_k|a_i\rangle\langle a_i|\hat{\rho}|a_j\rangle| = -1 + \sum_{i,j} |\langle a_i|\hat{\rho}|a_j\rangle| = C_{\ell_1}(\hat{\rho}; \{|a_i\rangle\}). \tag{120}$$

9.5. KD-positivity witnesses, uncertainty and (complete) incompatibility

As pointed out in section 2.6, whether a KD distribution is positive or non-positive can be of operational importance. A witness of this property is the total non-positivity, defined in equation (114). A state $\hat{\rho}$ is KD-positive ($\hat{\rho} \in \mathcal{E}_{\mathrm{KD}+}$) if and only if $\mathcal{N}[Q_{i_1,\ldots,i_k}(\hat{\rho})] = 1$. Thus, $\mathcal{N}(Q_{i_1,\ldots,i_k}(\hat{\rho}))$ faithfully witnesses KD non-positivity, much as the Wigner negative volume faithfully witnesses Wigner non-positivity [284]. However, to compute $\mathcal{N}(Q_{i_1,\ldots,i_k}(\hat{\rho}))$, one must know the full KD distribution. One might therefore seek a witness whose computation requires less information. Below, we review such witnesses.

From a theoretical perspective, an interesting witness is a state's support uncertainty, which we now introduce and study further. We, again, restrict our analysis to the standard KD distribution with respect to orthonormal bases $\{|a_i\rangle\}$ and $\{|b_j\rangle\}$. Let $\sharp X$ denote the cardinality of the set X. Let $|\psi\rangle \in \mathcal{H}$ denote a pure state. We define as follows the support uncertainties of $|\psi\rangle$ in the bases $\{|a_i\rangle\}$ and $\{|b_j\rangle\}$:

$$n_{\mathcal{A}}(\psi) \equiv \sharp\{i \in [1,d] \mid \langle a_i|\psi\rangle \neq 0\}, \quad \text{and} \quad n_{\mathcal{B}}(\psi) \equiv \sharp\{j \in [1,d] \mid \langle b_j|\psi\rangle \neq 0\}. \tag{121}$$

To interpret these definitions, we imagine expanding $|\psi\rangle$ in the basis $\{|a_i\rangle\}$ $[\{|b_j\rangle\}]$. The number of non-vanishing expansion coefficients is $n_{\mathcal{A}}(\psi)$ $[n_{\mathcal{B}}(\psi)]$. $n_{\mathcal{A}}(\psi)$ and $n_{\mathcal{B}}(\psi)$ enable the construction of an inequality [286] that can be viewed as an uncertainty relation:

$$n_{\mathcal{A}}(\psi) n_{\mathcal{B}}(\psi) \geqslant M^{-2}, \quad \text{wherein} \quad M \equiv \max_{i,j} \{|\langle a_i|b_j\rangle|\}. \tag{122}$$

The more $|\psi\rangle$ is localised in the $\{|a_i\rangle\}$ basis, the less the state must be localised in the $\{|b_j\rangle\}$ basis, and vice versa.

Figure 14 illustrates the lower bound inequality (122) with three examples of mutually unbiased bases. The set of points $(n_{\mathcal{A}}(\psi), n_{\mathcal{B}}(\psi))$ in the $n_{\mathcal{A}}$-$n_{\mathcal{B}}$ plane maps out an uncertainty diagram for the two bases. It is known that, for any bases $\{|a_i\rangle\}$ and $\{|b_j\rangle\}$, any state $|\psi\rangle$ that saturates inequality (122) is KD-positive [33]. The reverse implication is not generally true, however [33]. If the bases are mutually unbiased ($M = m_{\mathcal{A},\mathcal{B}} = 1/\sqrt{d}$), nevertheless, saturating inequality (122) is equivalent to $|\psi\rangle$'s being KD-positive [137]. The three panels of figure 14 reflect this relation. As a result of this relation, for mutually

unbiased bases in prime dimension, the only pure KD-positive states are the bases' elements. For further results about the geometric structure of the set of KD-positive states, see section 9.2 and [35].

Let us further explore the links amongst incompatibility, uncertainty and KD non-positivity in the general case—when the bases are not necessarily mutually unbiased bases. A different, additive uncertainty relation is more pertinent than inequality (122). The support uncertainty [33] is defined by

$$n_{\mathcal{A},\mathcal{B}}(\psi) \equiv n_{\mathcal{A}}(\psi) + n_{\mathcal{B}}(\psi). \tag{123}$$

This quantity is linked to KD positivity through the following theorem [33].

Theorem 9.3. *Suppose that* $m_{\mathcal{A},\mathcal{B}} \equiv \min_{i,j}\{|\langle a_i|b_j\rangle|\} > 0$. *If* $|\psi\rangle\langle\psi| \in \mathcal{E}_{\mathrm{KD}+}$, *then* $n_{\mathcal{A},\mathcal{B}}(\psi) \leqslant d+1$.

The theorem implies an uncertainty relation: If the uncertainty $n_{\mathcal{A},\mathcal{B}}(\psi) > d+1$, then the KD distribution must be non-positive. The bound of Theorem 9.3 holds even when $m_{\mathcal{A},\mathcal{B}} = 0$, such that the matrix with elements $\langle a_i|b_j\rangle$ has zeroes, if there are not too many zeroes [33]. In the absence of constraints on the two bases, other than that none of the $|b_j\rangle$s are parallel to any of the $|a_i\rangle$s, the upper bound becomes $n_{\mathcal{A},\mathcal{B}}(\psi) \leqslant \frac{3}{2}d$ [32]. Intermediate estimates of the type $n_{\mathcal{A},\mathcal{B}}(\psi) \leqslant d+s$, with $1 \leqslant s \leqslant \frac{d}{2}$, were recently proven in [287] under suitable conditions on the structure of the transition matrix $\hat{U}$. [32] encompasses also scenarios in which the KD distribution is expressed in terms of projectors other than rank-1 projectors.

Theorem 9.3 implies that the support uncertainty is a KD-positivity witness for pure states: If $n_{\mathcal{A},\mathcal{B}}(\psi)$ is large, then $|\psi\rangle$ is not KD-positive. However, $n_{\mathcal{A},\mathcal{B}}(\psi)$ is not always a faithful witness: There may exist states $|\psi\rangle$ for which $n_{\mathcal{A},\mathcal{B}}(\psi) \leqslant d+1$ but that are nevertheless KD-non-positive. See the three panels of figure 14 for an example. The support uncertainty nevertheless has an advantage over the total non-positivity: limited information about the pure state $|\psi\rangle$ suffices to determine the support uncertainty. An extension of this result to mixed states appears in [288]. The extension involves the convex roof of the support uncertainty.

Theorem 9.3 states that pure KD-positive states have low support uncertainty $n_{\mathcal{A},\mathcal{B}}$: no pure KD-positive state can lie above the line $n_{\mathcal{A},\mathcal{B}} = d+1$. Figure 14 illustrates this observation. Further examples appear in [33, 34]. Theorem 9.3, as such, evokes quantum optics. There, the Glauber–Sudarshan function of pure states is positive only for coherent states, which minimise $(\Delta x)^2 + (\Delta p)^2$ (the sum of two conjugate quadratures' squared uncertainties). Unlike in quantum optics, however $d+1$ does not necessarily *minimise* the support uncertainty $n_{\mathcal{A},\mathcal{B}}$. Figure 14 evidences this fact. Nevertheless, the analogy with quantum optics can be sharpened through *complete incompatibility*, introduced in [33, 34], to which we now turn our attention.

Two observables are said to be incompatible when they do not commute. This notion is weak: In the context of equation (5), it means that at least one of the elements in $\{|a_i\rangle\}$ and one of the elements in $\{|b_j\rangle\}$ satisfy $0 < |\langle a_i|b_j\rangle| < 1$. This criterion is satisfied if the bases are not permutations of one another. A slightly stronger requirement is that $M_{\mathcal{A},\mathcal{B}} \equiv \max_{ij}\{|\langle a_i|b_j\rangle|\} < 1$, which follows from $m_{\mathcal{AB}} \equiv \min_{ij}\{|\langle a_i|b_j\rangle|\} > 0$. This latter statement guarantees that all $|a_i\rangle$s are distinct from all $|b_j\rangle$s. The latter statement also entails that, if a measurement of $\{|a_i\rangle\}$ yields the outcome a_i, then a subsequent measurement of $\{|b_j\rangle\}$ can yield any value b_j. The uncertainty in this outcome is maximal for mutually unbiased bases, for which $M_{\mathcal{A},\mathcal{B}} = m_{\mathcal{A},\mathcal{B}} = \frac{1}{\sqrt{d}}$ (see equation (116)). For this reason, mutually unbiased bases are sometimes called *maximally incompatible*. To introduce complete incompatibility of the bases $\{|a_i\rangle\}$ and $\{|b_j\rangle\}$, we proceed as follows.

We consider measurements of $\hat{\Pi}_{\mathcal{A}}(S)$ and $\hat{\Pi}_{\mathcal{B}}(S)$. We regard these measurements as coarse-grained measurements of the observables $\hat{A}$ and $\hat{B}$. For all subsets $S, T \subset \{1, 2, \ldots, d\}$, we let
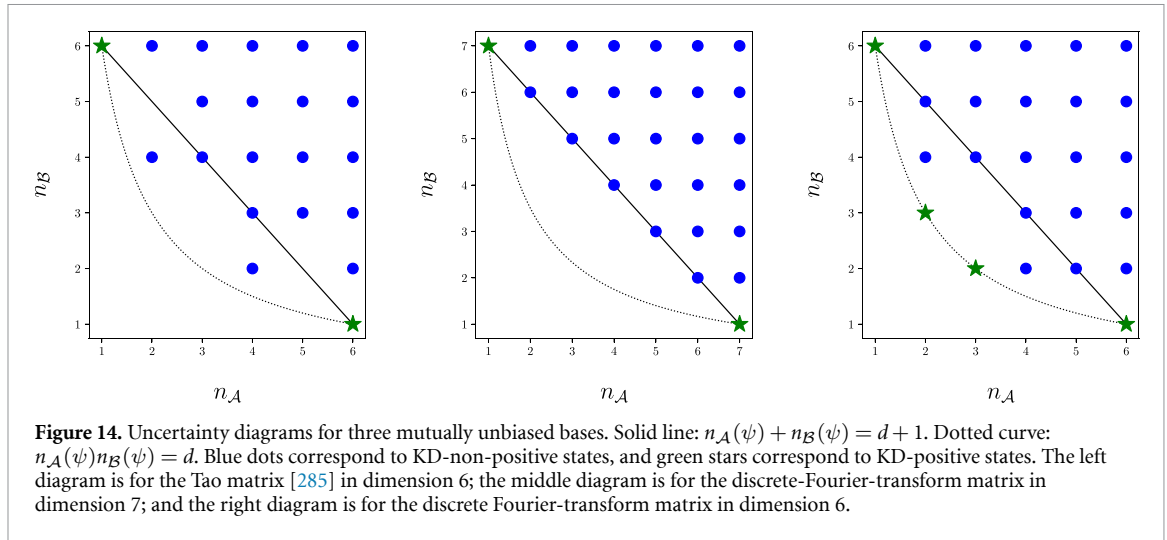
$$\hat{\Pi}_{\mathcal{A}}(S) = \sum_{i \in S} |a_i\rangle\langle a_i| \quad \text{and} \quad \hat{\Pi}_{\mathcal{B}}(T) = \sum_{j \in T} |b_j\rangle\langle b_j|. \tag{124}$$

In what follows, we shall designate by $\sharp S$ the number of elements belonging to S.

Suppose that a system is prepared in a state $|\psi\rangle$, a measurement of the projector $\hat{\Pi}_{\mathcal{A}}(S)$ yields the outcome 1, and a subsequent measurement of $\hat{\Pi}_{\mathcal{B}}(T)$ also yields 1. The system is then in the non-normalised state $\hat{\Pi}_{\mathcal{B}}(T)\hat{\Pi}_{\mathcal{A}}(S)|\psi\rangle$. Suppose, now, that a subsequent measurement of $\hat{\Pi}_{\mathcal{A}}(S)$ yields the outcome 1 with probability 1. In other words, suppose that the measurement of $\hat{\Pi}_{\mathcal{B}}(T)$ has not disturbed the outcome of the previous measurement of $\hat{\Pi}_{\mathcal{A}}(S)$. This condition is met only if $\hat{\Pi}_{\mathcal{B}}(T)\hat{\Pi}_{\mathcal{A}}(S)|\psi\rangle \in \hat{\Pi}_{\mathcal{A}}(S)\mathcal{H}$—in other words, provided that

$$\hat{\Pi}_{\mathcal{A}}(S)\mathcal{H} \cap \hat{\Pi}_{\mathcal{B}}(T)\mathcal{H} \neq \{0\}. \tag{125}$$

$\hat{\Pi}_{\mathcal{A}}(S)\mathcal{H}$ denotes the $\hat{\Pi}_{\mathcal{A}}(S)$ eigenspace associated with the eigenvalue 1. $\hat{\Pi}_{\mathcal{B}}(T)\mathcal{H}$ is defined similarly. Subsequent measurements of $\hat{\Pi}_{\mathcal{A}}(S)$ and $\hat{\Pi}_{\mathcal{B}}(T)$ will then consistently yield the outcome 1. Two successive

Figure 14. Uncertainty diagrams for three mutually unbiased bases. Solid line: $n_{\mathcal{A}}(\psi) + n_{\mathcal{B}}(\psi) = d + 1$. Dotted curve: $n_{\mathcal{A}}(\psi) n_{\mathcal{B}}(\psi) = d$. Blue dots correspond to KD-non-positive states, and green stars correspond to KD-positive states. The left diagram is for the Tao matrix [285] in dimension 6; the middle diagram is for the discrete-Fourier-transform matrix in dimension 7; and the right diagram is for the discrete Fourier-transform matrix in dimension 6.

measurements can be compatible in this way even if $\hat{\Pi}_{\mathcal{A}}(S)$ and $\hat{\Pi}_{\mathcal{B}}(T)$ do not commute; see [34] for examples.

Bases $\mathcal{A}$ and $\mathcal{B}$ are said to be completely incompatible when the above-described situation (equation (125)) is never realised [33, 34]. That is, $\mathcal{A}$ and $\mathcal{B}$ are completely compatible when, for all S, T with $\sharp S + \sharp T \leqslant d$,

$$\hat{\Pi}_{\mathcal{A}}(S)\mathcal{H} \cap \hat{\Pi}_{\mathcal{B}}(T)\mathcal{H} = \{0\}. \tag{126}$$

The restriction $\sharp S + \sharp T \leqslant d$ is needed since, whenever $\sharp S + \sharp T > d$, $\hat{\Pi}_{\mathcal{A}}(S)\mathcal{H} \cap \hat{\Pi}_{\mathcal{B}}(T)\mathcal{H} \neq \{0\}$ for dimensional reasons: Sufficiently coarse-grained measurements are always compatible in the above sense (equation (125)).

For completely incompatible bases, for all $|\psi\rangle \in \mathcal{H}$,

$$n_{\mathcal{A},\mathcal{B}}(\psi) \geqslant d + 1. \tag{127}$$

Therefore, if the bases $\mathcal{A}$ and $\mathcal{B}$ are completely incompatible and if $|\psi\rangle\langle\psi| \in \mathcal{E}_{\text{KD}+}$, then $n_{\mathcal{A},\mathcal{B}}(\psi) = d + 1$. In other words, when the bases are completely incompatible, the KD-positive states have minimal support uncertainties. The middle panel of figure 14 realises this situation, but the other two panels do not.

There is an open, dense set of bases that are completely incompatible [34]. However, it is not generally straightforward to determine if two given bases are completely incompatible. For example, consider two bases whose transition matrix is the discrete Fourier transform. These bases are completely incompatible only when the dimension d is prime [35]. Consequently, not all mutually unbiased bases are completely incompatible. In other words, *maximal* incompatibility does not imply *complete* incompatibility. Figure 14 illustrates this phenomenon. [287] proposes a further extension of the notion of complete incompatibility. In summary, the KD distribution can be a useful tool for designing and studying notions of incompatibility that extend beyond non-commutation.

10. Conclusion and outlook

Throughout this article, we have provided a comprehensive review of use cases of the KD distribution. In section 2, we defined the KD distribution and showed that it obeys a quasi-probabilistic version of Bayes' theorem. In section 3.1, we showed that non-real KD quasi-probabilities signal the disturbance of measurement-outcome probabilities. Similarly, as we outlined in section 3.2, one's ability to conduct quantum metrology, to learn unknown parameters encoded in a quantum state, hinges on non-real KD quasi-probabilities. In post-selected quantum metrology, one passes several quantum particles through a filter that distils their metrological information into the particles which pass the filter. In section 3.3, we showed that the rate of information distillation could be arbitrarily large if a KD distribution has negative components. In section 4, we reviewed weak values, pre- and post-selected observable averages of quantum states. We showed that non-positive KD distributions can lead to a weak value that lies outside the measured observable's spectrum. We also reviewed how such anomalous weak values can amplify metrological sensitivity to small unknown parameters. In section 5, we introduced the continuous-variable KD

distribution. Then, we reviewed how measurements of continuous-variable KD distributions have been used to directly measure quantum wavefunctions.

In classical thermodynamics, probability distributions describe statistic work and heat exchanges. In section 6, we showed how KD distributions can describe statistic exchanges in quantum thermodynamics. In section 7, we introduced the OTOC, a popular witness of many-body quantum chaos. First, we showed that the OTOC equals the average over a KD distribution. Second, whilst the OTOC struggles to distinguish information scrambling from decoherence, the KD distribution's non-positivity can witness scrambling more reliably.

We also summarised the KD distribution's importance in the foundations of quantum mechanics. In section 8.1, we showed how non-positive KD quasi-probabilities are required to violate Leggett–Garg inequalities (temporal Bell inequalities). In section 8.2, we described how the KD distribution is used in the consistent-histories interpretation of quantum mechanics. In section 8.3, we reviewed a rigorous notion of non-classicality: Contextuality. Non-positive KD quasi-probabilities can enable non-classical advantages in the operation of engines and weak measurements.

Given the KD distribution's diverse use cases, it is unsurprising that the distribution has been subject to growing mathematical research. We summarised, in section 9, the current knowledge about the KD distribution's mathematical properties. A basic necessary, but insufficient, condition for KD non-positivity is that $\mathcal{A} = \{|a_i\rangle\langle a_i|\}$ differ from $\mathcal{B} = \{|b_j\rangle\langle b_j|\}$. One could further ask, when is a quantum state $\hat{\rho}$ KD-positive? As outlined in section 9.2, there are several scenarios in which $\hat{\rho}$ is KD-positive if, and only if, $\hat{\rho}$ is a convex combination of $|a_i\rangle\langle a_i|$s and $|b_j\rangle\langle b_j|$s. For some pairs of bases, however, KD-positivity is not equivalent to this convex-combination property. Often, KD non-positivity signals non-classical advantages in quantum experiments. In section 9.3, we summarised properties of a KD distribution's total non-positivity. In section 9.5, we showed how to construct KD non-positivity witnesses that do not require full knowledge of the KD distribution. Also, we discussed how the KD distribution relates to a quantum state's uncertainties with respect to two observables.

Research on the KD distribution is ongoing. We conclude this review by listing a few promising outlook directions.

- **Quantum metrology:** we described the relationships between discrete-variable quantum metrology and non-positive KD distributions. Such relations have barely been explored in continuous-variable systems. Negative Wigner quasi-probabilities can signal advantages in continuous-variable metrology. An open question is whether the negativity requirement extends to KD representations of the experiments.
- **Weak values:** the KD distribution offers a means of understanding weak values' anomalous behaviours. A negative KD quasi-probability is required for the conditioned average of a weakly measured observable to lie outside the observable's spectrum. When is the KD distribution, or one of its extensions, the most appropriate quasi-probability distribution for characterising more-general conditioned sequences of weakened measurements? For example, as the weak-measurement strength increases, the KD quasi-probabilities will transform smoothly into projective-measurement probabilities. However, the intermediate-measurement-strength regime requires more study.
- **Direct measurement:** direct measurement provides an effective characterisation of quantum states, KD distributions, processes and detectors. An experimental opportunity is to capitalise on direct measurement's advantage over standard quantum tomography in applications to complex quantum systems. Possible targets include molecules and molecular processes, the multi-particle entangled states prepared by quantum circuits and quantum materials.
- **Quantum thermodynamics:** opportunities for future work include the marriage of KD distributions with conserved quantities (charges) that fail to commute with each other [289–291]. Such charges were overlooked for decades but engendered a growing subfield recently [175]. Example charges include the x-, y-, and z-components of spin [176, 292, 293]. The charges' noncommutation suggests them as observables whose eigenbases can define KD distributions naturally suited to quantum thermodynamics. Initial work on applying KD distributions to noncommuting thermodynamic charges has begun (sections 6 and 8.3.2) but merits expansion beyond currents.
- **Foundations of quantum mechanics:** one of the most rigorous notions of quantum phenomena is contextuality. If all quasi-probabilistic representation of an experiment are non-positive, the experiment is contextual [11]. Investigating infinitely many quasi-probability distributions is a formidable task. Could some distributions be more important than others? Often, the KD distribution is tailored to the operations that form an experiment. One could imagine that KD non-positivity may propagate to other quasi-probability representations.

● **Mathematical properties:** as reviewed above, the geometric structure of the convex set of KD-positive states is known in many cases. Nevertheless, for several operational tasks, we lack figures of merit for determining which KD-non-positive states are the most useful. Such metrics call for development.

Data availability statement

The data cannot be made publicly available upon publication because they are owned by a third party and the terms of use prevent public distribution. The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

ORCID iDs

Andrew N Jordan ⊙ https://orcid.org/0000-0002-9646-7013
Jeff S Lundeen ⊙ https://orcid.org/0000-0002-0978-2913
Nicole Yunger Halpern ⊙ https://orcid.org/0000-0001-8670-6212

References

[1] Dirac P A M 1942 *Proc. R. Soc. A* **180** 1
[2] Dirac P A M 1926 *Proc. R. Soc. A* **112** 661
[3] Dirac P A M 1926 *Proc. R. Soc. A* **110** 561
[4] Wigner E 1932 *Phys. Rev.* **40** 749
[5] Dirac P A M 1945 *Rev. Mod. Phys.* **17** 195
[6] Cohen L 1966 *Phil. Sci.* **33** 317
[7] Hudson R 1974 *Rep. Math. Phys.* **6** 249
[8] Srinivas M D and Wolf E 1975 *Phys. Rev. D* **11** 1477
[9] Griffiths R B 1984 *J. Stat. Phys.* **36** 219
[10] Hartle J B 2004 *Phys. Rev. A* **70** 022104
[11] Spekkens R W 2008 *Phys. Rev. Lett.* **101** 020401
[12] Allahverdyan A E 2015 *New J. Phys.* **17** 085005
[13] Kolmogorov A 1933 *Foundations of the Theory of Probability* ed N Morrison (Courier Dover Publications)
[14] Ville J 1948 *Câbles Transm.* **2** 61
[15] Moyal J E 1949 *Math. Proc. Camb. Phil. Soc.* **45** 99–124
[16] Wootters W K 1987 *Ann. Phys., NY* **176** 1
[17] Glauber R J 1963 *Phys. Rev.* **131** 2766
[18] Sudarshan E C G 1963 *Phys. Rev. Lett.* **10** 277
[19] Husimi K 1940 *Proc. Phys.-Math. Soc. Japan 3rd Ser.* **22** 264
[20] Mari A and Eisert J 2012 *Phys. Rev. Lett.* **109** 230503
[21] Booth R I, Chabaud U and Emeriau P-E 2022 *Phys. Rev. Lett.* **129** 230401
[22] Mandel L and Wolf E 1995 *Optical Coherence and Quantum Optics* (Cambridge University Press)
[23] Björk G, Klimov A B and Sanchez-Soto L L 2008 *Prog. Opt.* **51** 469
[24] Kirkwood J G 1933 *Phys. Rev.* **44** 31
[25] Terletsky Y P 1937 *J. Exp. Theor. Phys* **7** 1290
[26] Margenau H and Hill R N 1961 *Prog. Theor. Phys.* **26** 722
[27] Rihaczek A 1968 *IEEE Trans. Inf. Theory* **14** 369
[28] Johansen L M 2004 *Phys. Lett. A* **329** 184
[29] Johansen L M and Luis A 2004 *Phys. Rev. A* **70** 052115

[30] Yunger Halpern N 2017 *Phys. Rev.* A **95** 012120
[31] Allahverdyan A E 2014 *Phys. Rev.* E **90** 032137
[32] Arvidsson-Shukur D, Chevalier Drori J and Yunger Halpern N 2021 *J. Phys. A: Math. Theor.* **54** 284001
[33] De Bièvre S 2021 *Phys. Rev. Lett.* **127** 190404
[34] De Bièvre S 2023 *J. Math. Phys.* **64** 022202
[35] Langrenez C, Arvidsson-Shukur D R M and De Bièvre S 2024 *J. Math. Phys.* **65** 072201
[36] Langrenez C, Salmon W, Bièvre S D, Thio J J, Long C K and Arvidsson-Shukur D R M 2024 The set of kirkwood-dirac positive states is almost always minimal (arXiv:2405.17557 [quant-ph])
[37] Ferrie C 2011 *Rep. Prog. Phys.* **74** 116001
[38] Pashayan H, Wallman J J and Bartlett S D 2015 *Phys. Rev. Lett.* **115** 070501
[39] Schmid D, Baldijão R D, Ybng Y, Wagner R and Selby J H 2024 Kirkwood–Dirac representations beyond quantum states (and their relation to noncontextuality) (arXiv:2405.04573 [quant-ph])
[40] González Alonso J R, Yunger Halpern N and Dressel J 2019 *Phys. Rev. Lett.* **122** 040404
[41] Kohn J J and Nirenberg L 1965 *Commun. Pure Appl. Math.* **18** 269
[42] Steinberg A M 1995 *Phys. Rev.* A **52** 32
[43] Johansen L M 2007 *Phys. Rev.* A **76** 012119
[44] Hofmann H F 2012 *New J. Phys.* **14** 043031
[45] Yunger Halpern N, Swingle B and Dressel J 2018 *Phys. Rev.* A **97** 042105
[46] Hofmann H F 2014 *Phys. Rev.* A **89** 042115
[47] Bamber C and Lundeen J 2014 *Phys. Rev. Lett.* **112** 070405
[48] Arvidsson-Shukur D R, Yunger Halpern N, Lepage H V, Lasek A A, Barnes C H and Lloyd S 2020 *Nat. Commun.* **11** 3775
[49] Jenne J H and Arvidsson-Shukur D R M 2022 *Phys. Rev.* A **106** 042404
[50] Lupu-Gladstein N, Yilmaz Y B, Arvidsson-Shukur D R M, Brodutch A, Pang A O T, Steinberg A M and Yunger Halpern N 2022 *Phys. Rev. Lett.* **128** 220504
[51] Nielsen M A and Chuang I L 2011 *Quantum Computation and Quantum Information: 10th Anniversary Edition* 10th edn (Cambridge University Press)
[52] Helstrom C W 1976 *Quantum Detection and Estimation Theory* 1st edn (Academic) p 320
[53] Braunstein S L and Caves C M 1994 *Phys. Rev. Lett.* **72** 3439
[54] Cover T M and Thomas J A 2006 *Elements of Information Theory* 2nd edn (Wiley)
[55] Giovannetti V, Lloyd S and Maccone L 2006 *Phys. Rev. Lett.* **96** 010401
[56] Giovannetti V, Lloyd S and Maccone L 2011 *Nat. Photon.* **5** 222
[57] Krischek R, Schwemmer C, Wieczorek W, Weinfurter H, Hyllus P, Pezzé L and Smerzi A 2011 *Phys. Rev. Lett.* **107** 080504
[58] Demkowicz-Dobrzański R and Maccone L 2014 *Phys. Rev. Lett.* **113** 250801
[59] Smith J G, Barnes C H W and Arvidsson-Shukur D R M 2022 *Phys. Rev.* A **106** 062615
[60] Smith J G, Barnes C H W and Arvidsson-Shukur D R M 2024 *Phys. Rev.* A **109** 042412
[61] Salvati F, Salmon W, Barnes C H W and Arvidsson-Shukur D R M 2024 *Phys. Rev.* A **110** 022426
[62] Hance J R, Matsushita T and Hofmann H F 2024 *Quantum Sci. Technol.* **9** 045015
[63] Stone M H 1932 *Ann. Math.* **33** 643
[64] Dressel J and Jordan A N 2012 *Phys. Rev.* A **85** 012107
[65] Budiyono A, Agusta M K, Nurhandoko B E B and Dipojono H K 2023 *J. Phys. A: Math. Theor.* **56** 235304
[66] Budiyono A 2023 *Phys. Rev.* A **108** 012431
[67] Cramér H 2016 *Mathematical Methods of Statistics (PMS-9)* vol 9 (Princeton University Press)
[68] Rao C R 1992 *Breakthroughs in Statistics* (Springer) pp 235–47
[69] Hofmann H F 2011 *Phys. Rev.* A **83** 022106
[70] Tomm N *et al* 2021 *Nat. Nanotechnol.* **16** 399
[71] Dressel J, Malik M, Miatto F M, Jordan A N and Boyd R W 2014 *Rev. Mod. Phys.* **86** 307
[72] Harris J, Boyd R W and Lundeen J S 2017 *Phys. Rev. Lett.* **118** 070802
[73] Xu L, Liu Z, Datta A, Knee G C, Lundeen J S, Lu Y-q and Zhang L 2020 *Phys. Rev. Lett.* **125** 080501
[74] Arvidsson-Shukur D R M, McConnell A G and Yunger Halpern N 2023 *Phys. Rev. Lett.* **131** 150202
[75] Arvidsson-Shukur D R M, Barnes C H W, Lepage H V, Lloyd S, Lupu-Gladstein N, Salmon W and Salvati F 2024 Comment on "Saturating quantum advantages in postselected metrology with the positive Kirkwood–Dirac distribution" *Phys. Rev.* A **109** 066401
[76] Hosten O and Kwiat P 2008 *Science* **319** 787
[77] Dixon P B, Starling D J, Jordan A N and Howell J C 2009 *Phys. Rev. Lett.* **102** 173601
[78] Aharonov Y, Albert D Z and Vaidman L 1988 *Phys. Rev. Lett.* **60** 1351
[79] Duck I M, Stevenson P M and Sudarshan E C G 1989 *Phys. Rev.* D **40** 2112
[80] Tamir B and Cohen E 2013 *Quanta* **2** 7
[81] Von Neumann J 2018 *Mathematical Foundations of Quantum Mechanics: New Edition* vol 53 (Princeton University Press)
[82] Jordan A N and Siddiqi I A 2023 *Quantum Measurement: Theory and Practice* (Cambridge University Press)
[83] Dressel J 2015 *Phys. Rev.* A **91** 032116
[84] Viza G I, Martínez-Rincón J, Alves G B, Jordan A N and Howell J C 2015 *Phys. Rev.* A **92** 032127
[85] Starling D J, Dixon P B, Jordan A N and Howell J C 2009 *Phys. Rev.* A **80** 041803
[86] Jordan A N, Martínez-Rincón J and Howell J C 2014 *Phys. Rev.* X **4** 011031
[87] Feizpour A, Xing X and Steinberg A M 2011 *Phys. Rev. Lett.* **107** 133603
[88] Maga na-Loaiza O S, Mirhosseini M, Rodenburg B and Boyd R W 2014 *Phys. Rev. Lett.* **112** 200401
[89] Pang S, Alonso J R G, Brun T A and Jordan A N 2016 *Phys. Rev.* A **94** 012329
[90] Sinclair J, Hallaji M, Steinberg A M, Tollaksen J and Jordan A N 2017 *Phys. Rev.* A **96** 052128
[91] Starling D J, Dixon P B, Jordan A N and Howell J C 2010 *Phys. Rev.* A **82** 063822
[92] Dressel J, Lyons K, Jordan A N, Graham T M and Kwiat P G 2013 *Phys. Rev.* A **88** 023821
[93] Lyons K, Dressel J, Jordan A N, Howell J C and Kwiat P G 2015 *Phys. Rev. Lett.* **114** 170801
[94] Krafczyk C, Jordan A N, Goggin M E and Kwiat P G 2021 *Phys. Rev. Lett.* **126** 220801
[95] Wang Y-T *et al* 2016 *Phys. Rev. Lett.* **117** 230801
[96] Birkhoff G and von Neumann J 1936 *Ann. Math.* **37** 823

[97] Fano U 1957 *Rev. Mod. Phys.* **29** 74
[98] Pauli W 1980 *General Principles of Quantum Mechanics* (Springer)
[99] Vogel K and Risken H 1989 *Phys. Rev.* A **40** 2847
[100] Smithey D T, Beck M, Raymer M G and Faridani A 1993 *Phys. Rev. Lett.* **70** 1244
[101] Jordan A N 2016 *Physics* **9** 104
[102] Wootters W K and Zurek W H 1982 *Nature* **299** 802
[103] Arthurs E and Kelly J 1965 *Bell Syst. Tech. J.* **44** 725
[104] Leonhardt U and Paul H 1993 *Phys. Rev.* A **47** R2460
[105] Johansen L M and Mello P A 2008 *Phys. Lett.* A **372** 5760
[106] Lundeen J and Bamber C 2012 *Phys. Rev. Lett.* **108** 070402
[107] Thekkadath G, Hufnagel F and Lundeen J 2018 *New J. Phys.* **20** 113034
[108] Lundeen J S, Sutherland B, Patel A, Stewart C and Bamber C 2011 *Nature* **474** 188
[109] Lundeen J S and Resch K J 2005 *Phys. Lett.* A **334** 337
[110] Wu S and Mølmer K 2009 *Phys. Lett.* A **374** 34
[111] Salvail J, Agnew M, Johnson A, Bolduc E, Leach J and Boyd R 2013 *Nat. Photon.* **7** 316
[112] Thekkadath G, Giner L, Chalich Y, Horton M, Banker J and Lundeen J 2016 *Phys. Rev. Lett.* **117** 120401
[113] Kim Y, Kim Y-S, Lee S-Y, Han S-W, Moon S, Kim Y-H and Cho Y-W 2018 *Nat. Commun.* **9** 192
[114] Zhang C-R, Hu M-J, Hou Z-B, Tang J-F, Zhu J, Xiang G-Y, Li C-F, Guo G-C and Zhang Y-S 2020 *Phys. Rev.* A **101** 012119
[115] Malik M, Mirhosseini M, Lavery M, Leach J, Padgett M and Boyd R 2014 *Nat. Commun.* **5** 3115
[116] Bolduc E, Gariepy G and Leach J 2016 *Nat. Commun.* **7** 10439
[117] Ogawa K, Okazaki T, Kobayashi H, Nakanishi T and Tomita A 2021 *Opt. Express* **29** 19403
[118] Denkmayr T, Geppert H, Lemmel H, Waegell M, Dressel J, Hasegawa Y and Sponar S 2017 *Phys. Rev. Lett.* **118** 010402
[119] Lu D, Brodutch A, Li J, Li H and Laflamme R 2014 *New J. Phys.* **16** 053015
[120] Ozawa M 2011 *AIP Conf. Proc.* **1363** 53–62
[121] Yokota K, Yamamoto T, Koashi M and Imoto N 2009 *New J. Phys.* **11** 033011
[122] Lundeen J S and Steinberg A M 2009 *Phys. Rev. Lett.* **102** 020404
[123] Pan W-W *et al* 2019 *Phys. Rev. Lett.* **123** 150402
[124] Howland G, Lum D and Howell J 2014 *Opt. Express* **22** 18870
[125] Shi Z, Mirhosseini M, Margiewicz J, Malik M, Rivera F, Zhu Z and Boyd R 2015 *Optica* **2** 388
[126] Knarr S, Lum D, Schneeloch J and Howell J 2018 *Phys. Rev.* A **98** 023854
[127] Haapasalo E, Lahti P and Schultz J 2011 *Phys. Rev.* A **84** 052107
[128] Vallone G and Dequal D 2016 *Phys. Rev. Lett.* **116** 040502
[129] Cohen E and Pollak E 2018 *Phys. Rev.* A **98** 042112
[130] Calderaro L, Foletto G, Dequal D, Villoresi P and Vallone G 2018 *Phys. Rev. Lett.* **121** 230501
[131] Wen Y-L, Zhang S, Yan H and Zhu S-L 2022 *Chin. Phys.* B **31** 034206
[132] Hariri A, Curic D, Giner L and Lundeen J 2019 *Phys. Rev.* A **100** 032119
[133] Li Z, Xie L, Zhu J, Li W and Zhang Z 2021 *Appl. Phys. Express* **14** 102007
[134] Martinez-Becerril A, Bussières G, Curic D, Giner L, Abrahao R and Lundeen J 2021 *Quantum* **5** 599
[135] Yang M, Xiao Y, Liao Y-W, Liu Z-H, Xu X-Y, Xu J-S, Li C-F and Guo G-C 2020 *Laser Photon. Rev.* **14** 1900251
[136] Zhu J, Wang A, Liu X, Liu Y, Zhang Z and Gao F 2021 *Phys. Rev.* A **104** 032221
[137] Xu L, Xu H, Jiang T, Xu F, Zheng K, Wang B, Zhang A and Zhang L 2021 *Phys. Rev. Lett.* **127** 180401
[138] Zhang Y-X, Zhu X, Wu S and Chen Z-B 2017 *Ann. Phys., NY* **378** 13
[139] Gaikwad A, Singh G, Dorai K and Arvind 2023 *Eur. Phys. J.* D **77** 209
[140] Wen Y-L, Wang Y, Tian L-M, Zhang S, Li J, Du J-S, Yan H and Zhu S-L 2023 *Nat. Photon.* **17** 717–22
[141] Bollen V, Sua Y M and Lee K F 2010 *Phys. Rev.* A **81** 063826
[142] Hofmann H 2012 *Phys. Rev. Lett.* **109** 020408
[143] Buscemi F, Dall'Arno M, Ozawa M and Vedral V 2014 *Int. J. Quantum Inf.* **12** 1560002
[144] Thekkadath G, Saaltink R, Giner L and Lundeen J 2017 *Phys. Rev. Lett.* **119** 050405
[145] Lostaglio M, Belenchia A, Levy A, Hernández-Gómez S, Fabbri N and Gherardini S 2023 *Quantum* **7** 1128
[146] Wagner R, Schwartzman-Nowik Z, Paiva I L, Te'eni A, Ruiz-Molero A, Barbosa R S, Cohen E and Galvão E 2024 *Quantum Sci. Technol.* **9** 015030
[147] Buscemi F, Dall'Arno M, Ozawa M and Vedral V 2013 Direct observation of any two-point quantum correlation function (arXiv:1312.4240 [quant-ph])
[148] Chiribella G, Simonov K and Zhao X 2023 Dimension-independent weak value estimation via controlled swap operations (arXiv:2311.03941 [quant-ph])
[149] Parzygnat A J, Fullwood J, Buscemi F and Chiribella G 2024 *Phys. Rev. Lett.* **132** 110203
[150] Solinas P and Gasparinetti S 2016 *Phys. Rev.* A **94** 052103
[151] Miller H J D and Anders J 2017 *New J. Phys.* **19** 062001
[152] Levy A and Lostaglio M 2020 *PRX Quantum* **1** 010309
[153] Lostaglio M 2018 *Phys. Rev. Lett.* **120** 040602
[154] Lostaglio M, Belenchia A, Levy A, Hernández-Gómez S, Fabbri N and Gherardini S 2022 arXiv:2206.11783
[155] Teixidó-Bonfill A, Ortega A and Martín-Martínez E 2020 *Phys. Rev.* A **102** 052219
[156] Francica G 2022 *Phys. Rev.* E **105** 014101
[157] Francica G 2022 *Phys. Rev.* E **106** 054129
[158] Hernández-Gómez S, Gherardini S, Belenchia A, Lostaglio M, Levy A and Fabbri N 2024 *Phys. Rev. Res.* **6** 023280
[159] Díaz M G, Guarnieri G and Paternostro M 2020 *Entropy* **22** 1223
[160] Upadhyaya T, Braasch W F, Landi G T and Halpern N Y 2024 *PRX Quantum* **5** 030355
[161] Santini A, Solfanelli A, Gherardini S and Collura M 2023 *Phys. Rev.* B **108** 104308
[162] Pei J-H, Chen J-F and Quan H T 2023 *Phys. Rev.* E **108** 054109
[163] Gherardini S and De Chiara G 2024 *PRX Quantum* **5** 030201
[164] Jarzynski C 1997 *Phys. Rev.* E **56** 5018
[165] Crooks G E 1998 *J. Stat. Phys.* **90** 1481
[166] Collin D, Ritort F, Jarzynski C, Smith S B, Tinoco I and Bustamante C 2005 *Nature* **437** 231

[167] Liphardt J, Dumont S, Smith S B, Tinoco I and Bustamante C 2002 *Science* **296** 1832
[168] Pohorille A, Jarzynski C and Chipot C 2010 *J. Phys. Chem.* B **114** 10235
[169] Perarnau-Llobet M, Bäumer E, Hovhannisyan K V, Huber M and Acin A 2017 *Phys. Rev. Lett.* **118** 070601
[170] Lostaglio M 2020 *Phys. Rev. Lett.* **125** 230603
[171] Pratapsi S S, Deffner S and Gherardini S 2024 Quantum speed limit for kirkwood-dirac quasiprobabilities (arXiv:2402.07582 [quant-ph])
[172] Jarzynski C and Wójcik D K 2004 *Phys. Rev. Lett.* **92** 230602
[173] Jennings D and Rudolph T 2010 *Phys. Rev.* E **81** 061130
[174] Jevtic S, Rudolph T, Jennings D, Hirono Y, Nakayama S and Murao M 2015 *Phys. Rev.* E **92** 042113
[175] Majidy S, Braasch W F, Lasek A, Upadhyaya T, Kalev A and Yunger Halpern N 2023 *Nat. Rev. Phys.* **5** 689
[176] Yunger Halpern N, Beverland M E and Kalev A 2020 *Phys. Rev.* E **101** 042117
[177] Larkin A I and Ovchinnikov Y N 1969 *Sov. J. Exp. Theor. Phys.* **28** 1200
[178] Shenker S H and Stanford D 2014 *J. High Energy Phys.* JHEP03(2014)067
[179] Maldacena J, Stanford D and Yang Z 2017 *Fortschr. Phys.* **65** 1700034
[180] Yoshida B and Kitaev A 2017 arXiv:1710.03363
[181] Gao P, Jafferis D L and Wall A C 2017 *J. High Energy Phys.* JHEP12(2017)151
[182] Li J, Fan R, Wang H, Ye B, Zeng B, Zhai H, Peng X and Du J 2017 *Phys. Rev.* X **7** 031011
[183] Wei K X, Ramanathan C and Cappellaro P 2018 *Phys. Rev. Lett.* **120** 070501
[184] Nie X *et al* 2020 *Phys. Rev. Lett.* **124** 250601
[185] Niknam M, Santos L F and Cory D G 2020 *Phys. Rev. Res.* **2** 013200
[186] Britton J W, Sawyer B C, Keith A C, Wang C C J, Freericks J K, Uys H, Biercuk M J and Bollinger J J 2012 *Nature* **484** 489
[187] Islam R, Senko C, Campbell W C, Korenblit S, Smith J, Lee A, Edwards E E, Wang C-C J, Freericks J K and Monroe C 2013 *Science* **340** 583
[188] Gärttner M, Bohnet J G, Safavi-Naini A, Wall M L, Bollinger J J and Rey A M 2017 *Nat. Phys.* **13** 781
[189] Landsman K A, Figgatt C, Schuster T, Linke N M, Yoshida B, Yao N Y and Monroe C 2019 *Nature* **567** 61
[190] Joshi M K, Elben A, Vermersch B, Brydges T, Maier C, Zoller P, Blatt R and Roos C F 2020 *Phys. Rev. Lett.* **124** 240505
[191] Yan B, Moses S A, Gadway B, Covey J P, Hazzard K R A, Rey A M, Jin D S and Ye J 2013 *Nature* **501** 521
[192] Choi S *et al* 2017 *Nature* **543** 221
[193] Foxen B *et al* (Google AI Quantum) 2020 *Phys. Rev. Lett.* **125** 120504
[194] Chen B, Hou X, Zhou F, Qian P, Shen H and Xu N 2020 *Appl. Phys. Lett.* **116** 194002
[195] Pegahan S, Arakelyan I and Thomas J E 2021 *Phys. Rev. Lett.* **126** 070601
[196] Blok M S, Ramasesh V V, Schuster T, O'Brien K, Kreikebaum J M, Dahlen D, Morvan A, Yoshida B, Yao N Y and Siddiqi I 2021 *Phys. Rev.* X **11** 021010
[197] Garcia R J, Zhou Y and Jaffe A 2021 *Phys. Rev. Res.* **3** 033155
[198] Braumüller J *et al* 2022 *Nat. Phys.* **18** 172
[199] Mi X *et al* 2021 *Science* **374** 1479
[200] Geller M R, Arrasmith A, Holmes Z, Yan B, Coles P J and Sornborger A 2022 *Phys. Rev.* E **105** 035302
[201] Harris J, Yan B and Sinitsyn N A 2022 *Phys. Rev. Lett.* **129** 050602
[202] Blocher P D, Asaad S, Mourik V, Johnson M A I, Morello A and Mlmer K 2022 *Phys. Rev.* A **106** 042429
[203] Wang J-H *et al* 2022 *Phys. Rev. Res.* **4** 043141
[204] Green A M, Elben A, Alderete C H, Joshi L K, Nguyen N H, Zache T V, Zhu Y, Sundar B and Linke N M 2022 *Phys. Rev. Lett.* **128** 140601
[205] Dressel J, González Alonso J R, Waegell M and Yunger Halpern N 2018 *Phys. Rev.* A **98** 012132
[206] Mohseninia R, Alonso J R G and Dressel J 2019 *Phys. Rev.* A **100** 062336
[207] González Alonso J R, Shammah N, Ahmed S, Nori F and Dressel J 2022 arXiv:2201.08175
[208] Yunger Halpern N, Bartolotta A and Pollack J 2019 *Commun. Phys.* **2** 92
[209] Swingle B 2018 *Nat. Phys.* **14** 988
[210] Rozenbaum E B, Ganeshan S and Galitski V 2017 *Phys. Rev. Lett.* **118** 086801
[211] Talkner P, Lutz E and Hänggi P 2007 *Phys. Rev.* E **75** 050102
[212] Swingle B, Bentsen G, Schleier-Smith M and Hayden P 2016 *Phys. Rev.* A **94** 040302
[213] Yao N Y, Grusdt F, Swingle B, Lukin M D, Stamper-Kurn D M, Moore J E and Demler E A 2016 arXiv:1607.01801
[214] Zhu G, Hafezi M and Grover T 2016 *Phys. Rev.* A **94** 062329
[215] Zhang K and Wang J 2022 arXiv:2201.00385
[216] Zhang Y-L, Huang Y and Chen X 2019 *Phys. Rev.* B **99** 014303
[217] Zanardi P and Anand N 2021 *Phys. Rev.* A **103** 062214
[218] Touil A and Deffner S 2021 *PRX Quantum* **2** 010306
[219] Leggett A J and Garg A 1985 *Phys. Rev. Lett.* **54** 857
[220] Williams N S and Jordan A N 2008 *Phys. Rev. Lett.* **100** 026804
[221] Ruskov R, Korotkov A N and Mizel A 2006 *Phys. Rev. Lett.* **96** 200404
[222] Jordan A N, Korotkov A N and Büttiker M 2006 *Phys. Rev. Lett.* **97** 026805
[223] Emary C, Lambert N and Nori F 2013 *Rep. Prog. Phys.* **77** 016001
[224] Palacios-Laloy A, Mallet F, Nguyen F, Bertet P, Vion D, Esteve D and Korotkov A N 2010 *Nat. Phys.* **6** 442
[225] Goggin M E, Almeida M P, Barbieri M, Lanyon B P, O'Brien J L, White A G and Pryde G J 2011 *Proc. Natl Acad. Sci.* **108** 1256
[226] Knee G C *et al* 2012 *Nat. Commun.* **3** 606
[227] Groen J P, Ristè D, Tornberg L, Cramer J, de Groot P C, Picot T, Johansson G and DiCarlo L 2013 *Phys. Rev. Lett.* **111** 090506
[228] Bednorz A, Belzig W and Nitzan A 2012 *New J. Phys.* **14** 013009
[229] Dressel J, Broadbent C J, Howell J C and Jordan A N 2011 *Phys. Rev. Lett.* **106** 040402
[230] Lewis P J 2016 *Quantum Ontology: A Guide to the Metaphysics of Quantum Mechanics* (Oxford University Press)
[231] Everett H 1973 *The Theory of the universal Wavefunction* (Princeton University Press) pp 1–140
[232] Dewitt B S and Graham N 2015 *The Many-Worlds Interpretation of Quantum Mechanics* vol 63 (Princeton University Press)
[233] Bohm D 1952 *Phys. Rev.* **85** 166
[234] Faye J 2024 *The Stanford Encyclopedia of Philosophy* Summer 2024 edn, ed E N Zalta and U Nodelman (Metaphysics Research Lab, Stanford University)

[235] Gell-Mann M and Hartle J B 1993 *Phys. Rev.* D **47** 3345
[236] Omnes R 1999 *Understanding Quantum Mechanics* (Princeton University Press)
[237] Hartle J B 2007 *J. Phys. A: Math. Theor.* **40** 3101
[238] Hartle J B 2010 *Many Worlds? Everett, Quantum Theory, & Reality* ed S Saunders, J Barrett, A Kent and D Wallace (Oxford University Press)
[239] Hohenberg P C 2010 *Rev. Mod. Phys.* **82** 2835
[240] Zehnder L 1891 *Ein Neuer Interferenzrefraktor* (Springer)
[241] Mach L 1892 *Z. Instrum.* **12** 89
[242] Jennings D and Leifer M 2016 *Contemp. Phys.* **57** 60
[243] Brunner N, Cavalcanti D, Pironio S, Scarani V and Wehner S 2014 *Rev. Mod. Phys.* **86** 419
[244] Specker E 1960 *Dialectica* **14** 239
[245] Simon Kochen E S 1968 *Indiana Univ. Math. J.* **17** 59
[246] Budroni C, Cabello A, Gühne O, Kleinmann M and Larsson J-A 2022 *Rev. Mod. Phys.* **94** 045007
[247] Leifer M 2014 *Quanta* **3** 67
[248] Spekkens R W 2005 *Phys. Rev.* A **71** 052108
[249] Abramsky S and Brandenburger A 2011 *New J. Phys.* **13** 113036
[250] Peres A 1990 *Phys. Lett.* A **151** 107
[251] Mermin N D 1993 *Rev. Mod. Phys.* **65** 803
[252] Simon C, Brukner I C V and Zeilinger A 2001 *Phys. Rev. Lett.* **86** 4427
[253] Larsson J-Î 2002 *Europhys. Lett.* **58** 799
[254] Cabello A 2008 *Phys. Rev. Lett.* **101** 210401
[255] Klyachko A A, Can M A, Binicioğlu S and Shumovsky A S 2008 *Phys. Rev. Lett.* **101** 020403
[256] Hofmann H F 2015 *Phys. Rev.* A **91** 062123
[257] Hofmann H F 2020 *Phys. Rev.* A **102** 062215
[258] Ji M and Hofmann H F 2023 *Phys. Rev.* A **107** 022208
[259] Ji M and Hofmann H F 2024 *Quantum* **8** 1255
[260] Hance J R, Ji M and Hofmann H F 2023 *New J. Phys.* **25** 113028
[261] Einstein A *et al* 1905 *Ann. Phys., Lpz.* **17** 891
[262] Spekkens R W 2019 arXiv:1909.04628
[263] Leifer M S 2014 arXiv:1409.1570
[264] Beltrametti E and Bugajski S 1995 *J. Phys. A: Math. Gen.* **28** 3329
[265] Schmid D and Spekkens R W 2018 *Phys. Rev.* X **8** 011015
[266] Lostaglio M and Senno G 2020 *Quantum* **4** 258
[267] Kunjwal R, Lostaglio M and Pusey M F 2019 *Phys. Rev.* A **100** 042116
[268] Mazurek M D, Pusey M F, Kunjwal R, Resch K J and Spekkens R W 2016 *Nat. Commun.* **7** ncomms11780
[269] Mazurek M D, Pusey M F, Resch K J and Spekkens R W 2021 *PRX Quantum* **2** 020302
[270] Schmid D, Selby J H and Spekkens R W 2024 *Phys. Rev.* A **109** 022228
[271] Flatt K, Lee H, Carceller C R I, Brask J B and Bae J 2022 *PRX Quantum* **3** 030337
[272] Catani L, Leifer M, Scala G, Schmid D and Spekkens R W 2022 *Phys. Rev. Lett.* **129** 240401
[273] i Carceller C R, Flatt K, Lee H, Bae J and Brask J B 2022 *Phys. Rev. Lett.* **129** 050501
[274] Schmid D, Du H, Selby J H and Pusey M F 2022 *Phys. Rev. Lett.* **129** 120403
[275] Pusey M F 2014 *Phys. Rev. Lett.* **113** 200401
[276] Hiriart-Urruty J-B and Lemaréchal C 2001 *Fundamentals of Convex Analysis* (Springer) (https://doi.org/10.1007/978-3-642-56468-0)
[277] Designolle S, Farkas M and Kaniewski J 2019 *New J. Phys.* **21** 113053
[278] Thio J J, Salmon W, Barnes C H W, Bièvre S D and Arvidsson-Shukur D R M 2024 Contextuality can be verified with noncontextual experiments (arXiv:2412.00199 [quant-ph])
[279] Xu J 2022 Kirkwood–Dirac classical pure states (arXiv:2210.02876 [math-ph, physics:quant-ph])
[280] Durt T, Englert B-G, Bengtsson I and Życzkowski K 2010 *Int. J. Quantum Inf.* **8** 535
[281] Budiyono A and Dipojono H K 2023 *Phys. Rev.* A **107** 022408
[282] Budiyono A, Sumbowo J F, Agusta M K and Nurhandoko B E B 2023 Characterizing quantum coherence based on the nonclassicality of the Kirkwood–Dirac quasiprobability (arXiv:2309.09162 [quant-ph])
[283] Budiyono A, Gunara B E, Nurhandoko B E B and Dipojono H K 2023 *J. Phys. A: Math. Theor.* **56** 435301
[284] Kenfack A and Życzkowski K 2004 *J. Opt. B: Quantum Semiclass. Opt.* **6** 396
[285] Tao T 2004 *Math. Res. Lett.* **11** 251
[286] Donoho D L and Stark P B 1989 *SIAM J. Appl. Math.* **49** 906
[287] Xu J 2022 *Phys. Rev.* A **106** 022217
[288] Langrenez C, Bièvre S D and Arvidsson-Shukur D R M 2024 Convex roofs witnessing Kirkwood–Dirac nonpositivity (arXiv:2407.04558 [quant-ph])
[289] Lostaglio M, Jennings D and Rudolph T 2017 *New J. Phys.* **19** 043008
[290] Guryanova Y, Popescu S, Short A J, Silva R and Skrzypczyk P 2016 *Nat. Commun.* **7** 12049
[291] Yunger Halpern N, Faist P, Oppenheim J and Winter A 2016 *Nat. Commun.* **7** 12051
[292] Yunger Halpern N and Majidy S 2022 *npj Quantum Inf.* **8** 10
[293] Kranzl F, Lasek A, Joshi M K, Kalev A, Blatt R, Roos C F and Yunger Halpern N 2023 *PRX Quantum* **4** 020318