Quasi-Bell inequalities from symmetrized products of noncommuting qubit observables

Omar E. Gamel, and Graham R. Fleming

Citation: Journal of Mathematical Physics **58**, 052107 (2017); doi: 10.1063/1.4983918
View online: http://dx.doi.org/10.1063/1.4983918
View Table of Contents: http://aip.scitation.org/toc/jmp/58/5
Published by the American Institute of Physics

Articles you may be interested in

On the existence of impurity bound excitons in one-dimensional systems with zero range interactions
Journal of Mathematical Physics **58**, 052106 (2017); 10.1063/1.4983921

Uncertainty principles for Kac algebras
Journal of Mathematical Physics **58**, 052102 (2017); 10.1063/1.4983755

New concise upper bounds on quantum violation of general multipartite Bell inequalities
Journal of Mathematical Physics **58**, 052202 (2017); 10.1063/1.4982961

Evaluation of more general integrals involving universal associated Legendre polynomials
Journal of Mathematical Physics **58**, 052105 (2017); 10.1063/1.4983919

Fisher information and topological pressure
Journal of Mathematical Physics **58**, 052702 (2017); 10.1063/1.4983661

Curvature in Hamiltonian mechanics and the Einstein-Maxwell-dilaton action
Journal of Mathematical Physics **58**, 052901 (2017); 10.1063/1.4983665

JOURNAL OF MATHEMATICAL PHYSICS **58**, 052107 (2017)

Quasi-Bell inequalities from symmetrized products of noncommuting qubit observables

Omar E. Gamel[a)] and Graham R. Fleming
Department of Chemistry, University of California, Berkeley, California 94720, USA and Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

(Received 6 September 2016; accepted 8 May 2017; published online 26 May 2017)

Noncommuting observables cannot be simultaneously measured; however, under local hidden variable models, they must simultaneously hold premeasurement values, implying the existence of a joint probability distribution. We study the joint distributions of noncommuting observables on qubits, with possible criteria of positivity and the Fréchet bounds limiting the joint probabilities, concluding that the latter may be negative. We use symmetrization, justified heuristically and then more carefully via the Moyal characteristic function, to find the quantum operator corresponding to the product of noncommuting observables. This is then used to construct Quasi-Bell inequalities, Bell inequalities containing products of noncommuting observables, on two qubits. These inequalities place limits on the local hidden variable models that define joint probabilities for noncommuting observables. We find that the Quasi-Bell inequalities have a quantum to classical violation as high as $\frac{3}{2}$ on two qubit, higher than conventional Bell inequalities. The result demonstrates the theoretical importance of noncommutativity in the nonlocality of quantum mechanics and provides an insightful generalization of Bell inequalities. *Published by AIP Publishing.* [http://dx.doi.org/10.1063/1.4983918]

I. INTRODUCTION

Einstein, Podolsky, and Rosen posed a paradox in which they showed quantum mechanics leading to a superluminal interaction, that is, nonlocality, unless one assumes that it is an incomplete theory.[1] To preserve locality, their result suggested the existence of *local hidden variables*, inaccessible physical quantities shared between quantum systems that were prepared together. Bell showed with his well-known inequalities that local hidden variable models cannot reproduce all the predictions of quantum mechanics[2] demonstrating conclusively that quantum theory is at odds with local realism.

Clauser, Horne, Shimony, and Holt (CHSH) proposed a different Bell inequality for a bipartite system of two spin-$\frac{1}{2}$ particles[3] (i.e., two qubits), most commonly used today to demonstrate quantum nonlocality theoretically and experimentally.[4–9] It shows that a quantum expectation value violates a bound set by local realism by a factor of $\sqrt{2}$. Further generalizations of the CHSH inequality have been proposed[10–12] and extended to higher Hilbert space dimensions.[13,14] But the $\sqrt{2}$ violation for the CHSH spin-$\frac{1}{2}$ case increased only marginally.[15,16]

In the construction of Bell inequalities, one assumes the independent existence of premeasurement values. However, standard Bell inequalities only consider products of commuting observables, that is, observables on separate qubits. Although the premeasurement values of noncommuting observables (on the same qubit) are assumed to simultaneously exist, they are never multiplied together. The reason for this is that the axioms of quantum mechanics do not clearly determine the Hermitian

[a)]Electronic mail: ogamel@berkeley.edu

operator corresponding to a product of noncommuting observables. Thus, one runs into ambiguity when calculating the quantum analogue of the classical expression in the inequality. Some may object to the existence of such an operator, since quantum theory precludes the simultaneous measurement of noncommuting observables. This objection fades when we recall that *sums* of noncommuting observables clearly have operators, without implying simultaneous measurement, and it is possible that the same is true of products.

In this work, we explore the inclusion of products of noncommuting observables in Bell inequalities. Such products were examined by Arnault, although in the very different context of non-Hermitian observables.[17] Our goal is a deeper understanding of noncommutativity that yields insights into hidden variables, a relationship previously addressed by Fine.[18] In quantizing the inequalities, including the aforementioned products, we apply the symmetrization procedure, where the quantum operator for the product of noncommuting observables is the average of all possible permutations of the ordered product. The implication is that the expectation value of the classical product in a hidden variable model becomes the expectation value of the symmetrized operator product in the quantum model.

We justify symmetrization heuristically, and also more carefully via Moyal quantization.[19] On application of symmetrization, quantum theory violates our Quasi-Bell inequalities by a factor of $\frac{3}{2}$, larger than analogous violations of traditional Bell inequalities. The "quasi" prefix indicates that the inequalities contain products of noncommuting observables, which has important consequences in limiting the experimental verification that will be discussed.

We apply the symmetrization in the context of a spin-$\frac{1}{2}$ system with bivalent observables, i.e., where measurements yield ± 1, as in Ref. 20. This is in contrast to the usual context in which symmetrization is applied, that of the continuous conjugate variables of position (x) and momentum (p). Before the main results, we devote a great deal of attention to the joint probabilities of hidden variable models of noncommuting observables on the same spin-$\frac{1}{2}$ system. These set the context for the main results of the paper.

The remainder of the paper is divided into two main parts. The first part, up to and including Sec. V, analyzes the local hidden variable models that include joint probabilities, which provide the context for and build up to the symmetrization procedure. We begin with a review of quantum and local hidden variable models for commuting observables on two spin-$\frac{1}{2}$ particles in Sec. II. We then shift our discussion to noncommuting observables on one of the particles in Sec. III, where we outline limits placed by positivity and the Fréchet bounds on the joint probability distribution of two such observables. This is extended to three noncommuting observables in Sec. IV, since this is the minimum number needed for our Quasi-Bell inequality. These two sections take positivity and the Fréchet bounds as far as they will go, showing that they each lead to a trivial independent joint probability function. This suggests that we accept joint probabilities that are potentially negative, to get a more meaningful distribution. In this context, we introduce and justify the symmetrization procedure in Sec. V.

The second and more important part of the paper constructs the Quasi-Bell inequalities and finds their quantum violations. Section VI reviews the CHSH inequality and introduces a quasi-Bell inequality with the product of three noncommuting observables. Using the results of Secs. IV and V, the quantum to classical violation factor is found to be $\frac{3}{2}$. Higher order inequalities with more observables are constructed in Sec. VII. They are found to not increase the violation any further. In Sec. VIII, we discuss the relations these inequalities have to conventional Bell inequalities and hidden variable models on Werner States, and the implied ranges of locality and nonlocality.[21–24] We conclude with a discussion of the validity, conceptual strength, and limitations of the paper's results in Sec. IX.

We use the labels spin-$\frac{1}{2}$ particle and qubit interchangeably, as the two are essentially equivalent. Our overall system contains two qubits, one each for Alice and Bob. Commuting observables are those on separate qubits, and non-commuting observables are those on the same qubit. The Pauli matrices, $\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$, and $\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, together form a vector of matrices $\vec{\sigma}$. We index multiple noncommuting observables on the same qubit starting at 0 to facilitate generalization of our Quasi-Bell inequalities in Sec. VII.

II. COMMUTING OBSERVABLES

We begin by reviewing both the quantum and local (classical) probability distribution functions for two qubits. We highlight the relationship between the two models, and the implied joint probabilities.

A. Quantum probabilities

Suppose Alice and Bob share a bipartite system of two spin-$\frac{1}{2}$ particles. They each choose the direction along which to measure spin, $\hat{a}$, $\hat{b}$ yielding results $a, b \in \{1, -1\}$, respectively. Their choices of measurement direction are independent of one another. Since Alice and Bob have separate subsystems, quantum measurements on them commute with one another.

Assume Alice and Bob share a bipartite quantum state with density matrix ρ whose Bloch matrix components are the real vectors $\vec{u}, \vec{v}$ and matrix R. That is,

$$\rho = \frac{1}{4}\left(I \otimes I + \sum_i u_i \sigma_i \otimes I + \sum_j v_j I \otimes \sigma_j + \sum_{ij} R_{ij} \sigma_i \otimes \sigma_j\right).$$

The positivity of ρ places certain conditions on $\vec{u}, \vec{v}, R$.[25] The *quantum* joint probability is then given by

$$p_q(a, b|\hat{a}, \hat{b}) = \mathrm{Tr}\left[\rho P_a(\hat{a}) \otimes P_b(\hat{b})\right] = \frac{1}{4}\left(1 + a\,\hat{a} \cdot \vec{u} + b\,\hat{b} \cdot \vec{v} + ab\,\hat{a}^\dagger R \hat{b}\right), \tag{1}$$

where $P_a(\hat{a}) \equiv \frac{1}{2}(I + a\,\hat{a} \cdot \vec{\sigma})$ is a projection operator.

Individual subsystem probabilities may easily be calculated from (1) as $p_q(a|\hat{a}) = \frac{1}{2}\left(1 + a\,\hat{a} \cdot \vec{u}\right)$ and $p_q(b|\hat{b}) = \frac{1}{2}\left(1 + b\,\hat{b} \cdot \vec{v}\right)$. The quantum expectation values satisfy $\langle a \rangle_q = \hat{a} \cdot \vec{u}$, $\langle b \rangle_q = \hat{b} \cdot \vec{v}$, and $\langle ab \rangle_q = \hat{a}^\dagger R \hat{b}$.

B. Local hidden variables

If local hidden variables can describe the correlation between the two parties, then the *local* joint probability is

$$p_l(a, b|\hat{a}, \hat{b}) = \int p(\lambda) p_A(a|\hat{a}, \lambda) p_B(b|\hat{b}, \lambda) d\lambda, \tag{2}$$

where λ indicates the hidden variables, their distribution $p(\lambda)$ satisfying $\int p(\lambda) d\lambda = 1$, and $0 \leq p_A(a|\hat{a}, \lambda), p_B(b|\hat{b}, \lambda) \leq 1$ are the local probabilities of Alice and Bob, respectively.

If Alice were measuring her subsystem alone, her own local probability would be

$$p_A(a|\hat{a}) = \int p(\lambda) p_A(a|\hat{a}, \lambda) d\lambda. \tag{3}$$

Note that p_A denotes two different but related probability functions, with inclusion of the arguments removing ambiguity. Without loss of generality, we can write

$$p_A(a|\hat{a}, \lambda) = \frac{1}{2}\left[1 + a f_A(\hat{a}, \lambda)\right], \tag{4}$$

for function f_A satisfying $-1 \leq f_A \leq 1$. The analogue holds for Bob's local probability with the notation changing accordingly.

One can then rewrite the local joint probability (2) as

$$p_l(a, b|\hat{a}, \hat{b}) = \frac{1}{4}\left(1 + a\overline{f_A(\hat{a})} + b\overline{f_B(\hat{b})} + ab\overline{f_A(\hat{a}) f_B(\hat{b})}\right), \tag{5}$$

where the overline indicates the weighted average over the hidden variable λ, as per $\overline{f(\hat{n})} \equiv \int p(\lambda) f(\hat{n}, \lambda) d\lambda$.

We are now in a position to demand that the quantum probabilities be simulable via a local hidden variable model. Doing so requires equality of joint probabilities (1) and (5), yielding

$$\hat{a} \cdot \vec{u} = \overline{f_A(\hat{a})}, \qquad \hat{b} \cdot \vec{v} = \overline{f_B(\hat{b})}, \qquad \hat{a}^\dagger R \hat{b} = \overline{f_A(\hat{a}) f_B(\hat{b})}. \tag{6}$$

Bell's famous inequalities demonstrate that for some quantum states (e.g., the singlet state $\vec{u} = \vec{v} = 0$, $R = -I$), hidden variables are impossible.[2] Characterizing the exact set of $\vec{u}, \vec{v}, R$ for which (6) has a solution for some f_A, f_B is equivalent to characterizing two-qubit states simulable by local hidden variable models. Currently, this problem only has partial solutions.[23,26]

III. NONCOMMUTING OBSERVABLES

A. Joint probability

Before discussing Bell inequalities with noncommuting observables, we must answer the following question: what can one say about joint probabilities of noncommuting observables? More precisely, suppose Alice makes one of the two measurements along the directions $\hat{a}_0$, $\hat{a}_1$ yielding results $a_0, a_1 \in \{1, -1\}$. In general, these are incompatible observables due to their noncommutativity and cannot be measured simultaneously, despite theoretical attempts at such a definition.[27,28] Thus the question of a joint probability function does not arise operationally in quantum theory.

However, if we attempt to describe the correlations in terms of hidden variables, that is, outcomes existing prior to measurement, then we should be able to find a joint probability function. It is of interest to study the properties of such a joint probability distribution if it could exist. In fact, the existence of such a joint distribution has been reported to be equivalent to Bell's inequalities holding.[18] The question has also been investigated for the continuous degrees of freedom position and momentum.[29] We extend this to the bivalent qubit observables at hand.

We seek a reasonable expression for $p_A(a_0, a_1|\hat{a}_0, \hat{a}_1)$, the joint probability that Alice's measurement would yield outcomes a_0, a_1, respectively, for measurements along the directions $\hat{a}_0$, $\hat{a}_1$. We require that the joint probability yield the correct marginal probabilities for each measurement. For example, $\sum_{a_1} p_A(a_0, a_1|\hat{a}_0, \hat{a}_1) = p_A(a_0|\hat{a}_0) = \frac{1}{2}\left(1 + a_0\, \hat{a}_0 \cdot \vec{u}\right)$. Making use of the known marginal probabilities and the probability function's completeness, we can write without loss of generality,

$$p_A(a_0, a_1|\hat{a}_0, \hat{a}_1) = \frac{1}{4}\left(1 + a_0\, \hat{a}_0 \cdot \vec{u} + a_1\, \hat{a}_1 \cdot \vec{u} + a_0 a_1 \langle a_0 a_1 \rangle\right). \tag{7}$$

The expectation value $\langle a_0 a_1 \rangle$ is some as yet undetermined function of $\hat{a}_0, \hat{a}_1$, and $\vec{u}$.

It is instructive to compare the noncommuting joint probability function (7) with commuting one (1). The two are very similar: with the expression for commuting observables potentially simulating that for noncommuting ones if we set $\vec{v} = \vec{u}$ and the expectation of the product takes the form $\langle a_0 a_1 \rangle = \hat{a}_0 R_{nc} \hat{a}_1$ for some R_{nc}, the "correlation matrix" for simulating the noncommuting measurements.

We now turn our attention to potential criteria to determine or limit the joint expectation value $\langle a_0 a_1 \rangle$. We consider two criteria that are natural in classical probability theory: positivity and satisfaction of the Fréchet inequalities for joint probabilities.

B. Positive distribution

One obvious condition is the *positivity* of the joint probability distribution,

$$p(a_0, a_1|\hat{a}_0, \hat{a}_1) \geq 0, \tag{8}$$

such that the joint probabilities are realizable non-negative values. We enforce (8) by requiring the right hand side of (7) to be non-negative for the four possible values of the pair a_0, a_1. This yields conditions that are instructively summarized in the following inequalities:

$$-(1 \pm \hat{a}_0 \cdot \vec{u})(1 \pm \hat{a}_1 \cdot \vec{u}) \leq \langle a_0 a_1 \rangle - (\hat{a}_0 \cdot \vec{u})(\hat{a}_1 \cdot \vec{u}) \leq (1 \pm \hat{a}_0 \cdot \vec{u})(1 \mp \hat{a}_1 \cdot \vec{u}), \tag{9}$$

which hold for both the upper and lower signs. Defining the difference quantity in the middle,

$$D(\hat{a}_0, \hat{a}_1) \equiv \langle a_0 a_1 \rangle - (\hat{a}_0 \cdot \vec{u})(\hat{a}_1 \cdot \vec{u}), \tag{10}$$

we seek its allowable functional forms. Since $D(\hat{a}_0, \hat{a}_1)$ must be basis independent, it must be a function of dot products of $\hat{a}_0, \hat{a}_1$ and $\vec{u}$.

1. Pure state

In case the underlying single qubit state is pure $|\vec{u}| = 1$, the minimum of the upper bound and the maximum of the lower bound are both zero, attained for $\hat{a}_0 = \pm\vec{u}$ or $\hat{a}_1 = \pm\vec{u}$. More precisely, $D(\pm\vec{u}, \hat{a}_1) = 0$, $\forall\, \hat{a}_1$ and $D(\hat{a}_0, \pm\vec{u}) = 0$, $\forall\, \hat{a}_0$. It can be shown that this implies $D(\hat{a}_0, \hat{a}_1)$ which is identically zero, and therefore

$$\langle a_0 a_1 \rangle \equiv (\hat{a}_0 \cdot \vec{u})(\hat{a}_1 \cdot \vec{u}). \tag{11}$$

That is, the expectation value of the product is equal to the product of expectation values, meaning the two measurements are *independent*. We can use commuting measurements on two qubits to simulate these noncommuting measurements on a single qubit if we set the correlation matrix as an outer product $R_{nc} = \vec{u}\vec{u}^{\dagger}$, i.e., the two qubits are in a product state. This is not surprising, since we required positivity, and the only physical two-qubit states where each individual qubit's state is pure are product states.

However, given that the two noncommuting measurements are on the same qubit, this independence is highly unexpected. This is partial because in the limit $\hat{a}_1 \to \hat{a}_0$ we get $\langle a_0 a_1 \rangle \to (\hat{a}_0 \cdot \vec{u})^2$, which is not identically unity as $\langle a_0^2 \rangle = 1$ would imply. It is intuitively expected that different measurements of the same qubit should be at the very least correlated. If this intuition is correct, the independence derived above casts doubt on the assumption of positivity of the joint probability.

In a different context, Ballentine found independent joint probability distributions of noncommuting observables to satisfy positivity, but dismissed this on physical grounds.[29] If we follow suit and choose to reject independence of noncommuting measurements, we will have to accept negative probabilities, a recurring theme within quantum theory, which we address in more detail in Sec. V.

2. Mixed state

In case the underlying single qubit state is mixed, $u \equiv |\vec{u}| < 1$, there is an allowed range for $D(\hat{a}_0, \hat{a}_1)$ and hence $\langle a_0 a_1 \rangle$. In particular, (9) implies

$$(-1 + \hat{a} \cdot \vec{u})(1 - u) \le D(\hat{a}, \hat{u}), D(\hat{u}, \hat{a}) \le (1 + \hat{a} \cdot \vec{u})(1 - u). \tag{12}$$

This condition will be satisfied by any $D = (\alpha + \hat{a} \cdot \vec{u})(1 - u)$ for some $-1 \le \alpha \le 1$. For example, we may define $D(\hat{a}_0, \hat{a}_1) \equiv \left[\hat{a}_0 \cdot \vec{u} + \hat{a}_1 \cdot \vec{u} + (1 - u)\hat{a}_0 \cdot \hat{a}_1\right](1 - u)$, where the two right most terms in the square brackets constitute α and encompass the range between 1 and −1. Note that this definition of $D(\hat{a}_0, \hat{a}_1)$ is symmetric in its two arguments, as one would expect. The expectation value of the product then takes the interesting form,

$$\langle a_0 a_1 \rangle \equiv \left[\hat{a}_0 \cdot \vec{u} + \hat{a}_1 \cdot \vec{u} + (1 - u)\hat{a}_0 \cdot \hat{a}_1\right](1 - u) + (\hat{a}_0 \cdot \vec{u})(\hat{a}_1 \cdot \vec{u}). \tag{13}$$

If the underlying qubit is maximally mixed, $\vec{u} = 0$, then (9) simplifies to the weak condition $-1 \le \langle a_0 a_1 \rangle \le 1$. In this case, there is a great deal of freedom in assigning a functional form to $\langle a_0 a_1 \rangle$ that always lies within this range. We may follow the form of (13), which yields $\langle a_0 a_1 \rangle \equiv \hat{a}_0 \cdot \hat{a}_1$ for $u = 0$ and incidentally corresponds to the symmetrization of noncommuting observables, the subject of Sec. V.

One can conclude that if the quantum state is mixed, requiring the joint probability of noncommuting observables to be positive does not necessarily imply independence. This is interesting in its own right and opens the door for reasonable and positive joint probabilities for some quantum states. However, the implied independence for pure states means that one cannot demand positivity in general.

C. Fréchet inequalities

Possible values for the joint probability of two classical events are bounded by the individual (marginal) probability of each event. For example, the joint probability of two events each with probability unity (zero) must itself be unity (zero). The joint probability of two events each with probability $\frac{1}{2}$ may be 0 if they are mutually exclusive, $\frac{1}{4}$ if they are independent, $\frac{1}{2}$ if they fully coincide, or any value in between.

More precisely, classical joint probabilities must satisfy the Fréchet inequalities,[30,31] which place bounds based on the individual probabilities,

$$p(a_0|\hat{a}_0) + p(a_1|\hat{a}_1) - 1 \le p(a_0, a_1|\hat{a}_0, \hat{a}_1) \le \min\{p(a_0|\hat{a}_0), p(a_1|\hat{a}_1)\}. \tag{14}$$

Plugging the marginal and joint probabilities implied by (7) into (14), rearranging and simplifying yield the following two inequalities:

$$\frac{1}{4}\left(1 - a_0\,\hat{a}_0\cdot\vec{u} - a_1\,\hat{a}_1\cdot\vec{u} + a_0a_1\,\langle a_0a_1\rangle\right) \ge 0, \tag{15}$$

$$\frac{1}{4}\left(1 + \min\{a_0\,\hat{a}_0\cdot\vec{u}, a_1\,\hat{a}_1\cdot\vec{u}\} - \max\{a_0\,\hat{a}_0\cdot\vec{u}, a_1\,\hat{a}_1\cdot\vec{u}\} - a_0a_1\,\langle a_0a_1\rangle\right) \ge 0. \tag{16}$$

Upon comparing with (7), it is evident the left hand side of (15) is equal to $p(-a_0, -a_1|\hat{a}_0, \hat{a}_1)$. Similarly, the left hand side of (16) is equal to $p(-a_0, a_1|\hat{a}_0, \hat{a}_1)$ or $p(a_0, -a_1|\hat{a}_0, \hat{a}_1)$ depending on whether $a_0\,\hat{a}_0\cdot\vec{u}$ or $a_1\,\hat{a}_1\cdot\vec{u}$ is larger. Since all the above inequalities are meant to hold $\forall\, a_0, a_1 \in \{1, -1\}$, (15) and (16) are each equivalent to (8). In other words, *requiring the Fréchet inequalities hold is identical to the requirement of positivity of the joint probability*, which as we showed in Sec. III B, has the unwanted consequence of independence for pure states.

IV. THREE OR MORE OBSERVABLES

Suppose we increase the number of noncommuting observables, starting with three. A derivation similar to that of (7) will show that the triple joint probability, without loss of generality, can be written as

$$\begin{aligned} p(a_0, a_1, a_2|\hat{a}_0, \hat{a}_1, \hat{a}_2) = \frac{1}{8}\big(&1 + a_0\,\hat{a}_0\cdot\vec{u} + a_1\,\hat{a}_1\cdot\vec{u} + a_2\,\hat{a}_2\cdot\vec{u} + a_0a_1\,\langle a_0a_1\rangle \\ &+ a_0a_2\,\langle a_0a_2\rangle + a_1a_2\,\langle a_1a_2\rangle + a_0a_1a_2\,\langle a_0a_1a_2\rangle\big). \end{aligned} \tag{17}$$

If we require this joint probability to always be nonnegative, we may engage in a derivation similar to that of inequality (9). For pure states $|\vec{u}| = 1$, the result will be analogous to the two-observable case. That is,

$$\langle a_0a_1a_2\rangle \equiv (\hat{a}_0\cdot\vec{u})(\hat{a}_1\cdot\vec{u})(\hat{a}_2\cdot\vec{u}). \tag{18}$$

Thus, requiring positivity of the joint probability distribution of three noncommuting observables for a pure state also implies their independence. The same procedure can be extended to show the independence of any number of noncommuting observables on a pure state, if positivity is required.

However, it is interesting that the two Fréchet inequalities when applied to three (or more) noncommuting observables are not both equal to the positivity condition, as was the case for two observables. Given three observables, the two-observable Fréchet inequalities (14) will still hold for all pairs. The additional three-observable Fréchet inequalities are

$$\max_{ijk}\{p(a_i|\hat{a}_i) + p(a_j, a_k|\hat{a}_j, \hat{a}_k) - 1\} \le p(a_0, a_1, a_2|\hat{a}_0, \hat{a}_1, \hat{a}_2) \le \min_{mn}\{p(a_m, a_n|\hat{a}_m, \hat{a}_n)\}, \tag{19}$$

where the indices i, j, k are distinct, l, m, n are distinct, and all take values 0, 1, 2. Note that (19) is obtained from (14) by treating the occurrence $(a_0, a_1, a_2|\hat{a}_0, \hat{a}_1, \hat{a}_2)$ as a two-way conjunction of $(a_i|\hat{a}_i)$ and $(a_j, a_k|\hat{a}_j, \hat{a}_k)$. We could also treat it as a three-way conjunction of $(a_i|\hat{a}_i)$, $(a_j|\hat{a}_j)$, and $(a_k|\hat{a}_k)$. However, this more reductionist approach yields a weaker inequality that is implied by (19) and (14) anyway.

We proceed by plugging (17) and (7) into (19), rearranging, and simplifying. This yields the following two inequalities for the lower and upper bounds, respectively:

$$\frac{1}{8}\left(3 - 3a_i\,\hat{a}_i\cdot\vec{u} - a_j\,\hat{a}_j\cdot\vec{u} - a_k\,\hat{a}_k\cdot\vec{u} + a_ia_j\,\langle a_ia_j\rangle + a_ia_k\,\langle a_ia_k\rangle - a_ja_k\,\langle a_ja_k\rangle + a_ia_ja_k\,\langle a_ia_ja_k\rangle\right) \ge 0, \tag{20}$$

$$\frac{1}{8}\left(1 - a_l\,\hat{a}_l\cdot\vec{u} + a_m\,\hat{a}_m\cdot\vec{u} + a_n\,\hat{a}_n\cdot\vec{u} - a_la_m\langle a_la_m\rangle - a_la_n\langle a_la_n\rangle + a_ma_n\langle a_ma_n\rangle - a_la_ma_n\langle a_la_ma_n\rangle\right) \ge 0, \tag{21}$$

where the indices i, j, k and l, m, n are assumed to be the ones satisfying the maximum/minimum in (19). These two bounds can be rewritten as

$$p(-a_i, -a_j, -a_k|\hat{a}_i, \hat{a}_j, \hat{a}_k) + p(-a_i, -a_j, a_k|\hat{a}_i, \hat{a}_j, \hat{a}_k) + p(-a_i, a_j, -a_k|\hat{a}_i, \hat{a}_j, \hat{a}_k) \geq 0, \tag{22}$$

$$p(-a_l, a_m, a_n|\hat{a}_l, \hat{a}_m, \hat{a}_n) \geq 0. \tag{23}$$

The lower Fréchet bound (22) is implied by, but weaker than the positivity condition. The upper bound (23) is equivalent to the positivity condition.

Therefore, we may conclude that for the three observables, the Fréchet bounds, taken together, are again equivalent to the positivity condition. This seems to hold for more observables as well.

V. SYMMETRIZATION

Thus far, we considered requiring the joint probabilities to be nonnegative or, equivalently, that they satisfy the Fréchet inequalities of classical probability theory. It was found that in the case of pure quantum states, these imply the independence of measurements of any noncommuting observables. We deemed this independence unsatisfactory on physical grounds and now seek a different approach. It is obvious that any approach that does not explicitly require positivity of the joint distribution may yield some negative probabilities. Though physically meaningless, negative probabilities need not be an operational problem since they correspond to impossible simultaneous measurements.[32] In Feynman's words, they may be used if "the situation for which the probability appears to be negative is not one that can be verified directly."[33]

The negativity of joint probabilities of noncommuting observables is well known.[28,34–37] It has even been shown as equivalent to fundamental nonclassical features of quantum theory.[38–42] This phenomenon is also related to the Wigner quasi-probability distribution taking negative values.[43,44]

As per (7), thus, we need an expression for $\langle a_0 a_1 \rangle$ based on the quantum expectation value. At first glance, the product of noncommuting observables which cannot be simultaneously measured may seem precluded by quantum mechanics. However, consider that the sum of noncommuting observables is a well-defined observable in its own right, e.g., $\sigma_1 + \sigma_2$. As Bell argued, "a measurement of a sum of noncommuting observables cannot be made by combining trivially the results of separate observations on the two terms — it requires a quite distinct experiment."[45] Similarly, the product does not involve measuring each operator and multiplying the results, rather the product is itself a legitimate observable whose operator can be derived in a manner consistent with quantization rules. The question is then finding such an operator.

Some authors have explored joint quasiprobability distributions for spin-$\frac{1}{2}$ states.[46–48] As expected, the quasiprobabilities may be negative. However, some of these methods lack symmetry in the arguments, or basis independence. Others do not easily lend themselves to variable directions of the spin operators. Therefore we pursue a more flexible approach.

A. Heuristic derivation

Commonly used heuristic arguments consistent with quantum theory provide *symmetrization* as the quantization of the product of two noncommuting observables, such as position and momentum. That is, the quantization of the product is set equal to the average of all possible permutations of the product of the quantizations. The result is Hermitian, symmetric in all the operators, and reduces to the simple product if the operators commute. For example, $xp \to \frac{1}{2}(\hat{x}\hat{p} + \hat{p}\hat{x})$. Many authors have made use of such a symmetrization.[20,49–52] Some ambiguity arises if any of the quantities in the product are raised to a power greater than unity. However, this is not a concern here.

Note that there are two related but subtly different concepts here, both involving a classical quantity in some way underlying a quantum operator. One is when a classical expression is *quantized* into a quantum observable, and the other is when a classical hidden variable is posited to model the measurement results on the said observable. It is in the former case that symmetrization is traditionally used, while here we use it in the latter, as others did.[20,35]

Going forward, we now find the quantum operator that corresponds to the classical product $a_0 a_1$, with the intention of applying standard quantum measurements to get the joint probability function. The quantizations of a_0, a_1 are $\hat{a}_0 \cdot \vec{\sigma}$, $\hat{a}_1 \cdot \vec{\sigma}$, respectively. Applying symmetrization for two observables, the result is simply half the anticommutator, yielding

$$a_0 a_1 \to \frac{1}{2}\left\{\hat{a}_0 \cdot \vec{\sigma}, \hat{a}_1 \cdot \vec{\sigma}\right\} = \hat{a}_0 \cdot \hat{a}_1 I. \tag{24}$$

The symmetrized observable is proportional to the identity matrix, with two identical eigenvalues $\hat{a}_0 \cdot \hat{a}_1$, not ± 1 as the individual realizations of the classical hidden variable product $a_0 a_1$. Nonetheless, there is no inconsistency. To see this, consider sums instead of products once more. The fact that the eigenvalues of $\sigma_1 + \sigma_2$ are not sums of eigenvalues of σ_1 and σ_2 does not preclude a local hidden variable model, whose individual realizations yield eigenvalues of σ_1 and σ_2, but not of their sum.

This is the essence of Bell's refutation of von Neumann's purported proof for the impossibility of hidden variables.[53] Bell argued that there is no reason to require the individual realizations of a hidden variable model to be additive over sums of noncommuting observables, since the observables cannot be measured simultaneously anyway. The function of a hidden variable is the model "to reproduce the *measurable* peculiarities of quantum mechanics when *averaged over*,"[45] that is, to reproduce the expectation values. Similar reasoning may be applied to products of noncommuting observables.

To be fully consistent with the statistical predictions of quantum mechanics, a_0, a_1 of the hidden variable model must be correlated with one another in such a way that the classical expectation value of their product coincides with the quantum expectation value of its quantization, which as in (24), is the symmetrized product operator. Aside however, note that the analogy of the sum to product breaks down in that the expectation value of sum is the sum of expectation values, but the expectation value of the product is not related to the individual expectation values in a simple way.

Moving forward, we find that the expectation value of the product is then

$$\langle a_0 a_1 \rangle = \mathrm{Tr}\left[\rho_A \hat{a}_0 \cdot \hat{a}_1 I\right] = \hat{a}_0 \cdot \hat{a}_1, \tag{25}$$

which very interestingly is independent of Alice's reduced (single qubit) quantum state $\rho_A \equiv \mathrm{Tr}_B[\rho]$ and its Bloch vector $\vec{u}$.

The joint probability associated with symmetrization is then

$$\begin{aligned} p(a_0, a_1 | \hat{a}_0, \hat{a}_1) &= \mathrm{Tr}\left[\rho_A \frac{1}{2}\left\{P_{a_0}(\hat{a}_0), P_{a_1}(\hat{a}_1)\right\}\right] \\ &= \frac{1}{4}\left(1 + a_0\, \hat{a}_0 \cdot \vec{u} + a_1\, \hat{a}_1 \cdot \vec{u} + a_0 a_1\, \hat{a}_0 \cdot \hat{a}_1\right). \end{aligned} \tag{26}$$

As expected, the joint probability in (26) is sometimes negative. For example, if $a_0 = a_1 = -1$, $\hat{a}_0 = (1, 0, 0)$, $\hat{a}_1 = (0, 1, 0)$, $\vec{u} = (1, 1, 0)/\sqrt{2}$, then it yields $p(a_0, a_1 | \hat{a}_0, \hat{a}_1) = (1 - \sqrt{2})/4 = -0.104$. We reiterate that the negative probability is unobservable directly since the qubit cannot be measured along $\hat{a}_0$ and $\hat{a}_1$ simultaneously and hence does not lead to operational contradictions.

Now turning our attention to the three noncommuting observables, we seek the quantum operator for the classical product $a_0 a_1 a_2$. Let S be the set of the six possible permutations of 0, 1, 2. Then symmetrization yields

$$a_0 a_1 a_2 \to \frac{1}{6} \sum_{lmn \in S} (\hat{a}_l \cdot \vec{\sigma})(\hat{a}_m \cdot \vec{\sigma})(\hat{a}_n \cdot \vec{\sigma}) = \vec{a}_{012} \cdot \vec{\sigma}, \tag{27}$$

where the symmetric product vector $\vec{a}_{012}$ is defined as

$$\vec{a}_{012} \equiv \frac{1}{3}\left[(\hat{a}_1 \cdot \hat{a}_2)\hat{a}_0 + (\hat{a}_2 \cdot \hat{a}_0)\hat{a}_1 + (\hat{a}_0 \cdot \hat{a}_1)\hat{a}_2\right], \tag{28}$$

fully symmetric in $\hat{a}_0$, $\hat{a}_1$, and $\hat{a}_2$ as required. Note that $|\vec{a}_{012}| \le 1$, and the eigenvalues of $\vec{a}_{012} \cdot \vec{\sigma}$ are $\pm|\vec{a}_{012}|$. This triple symmetrization has previously been applied by Barut *et al.*[20]

The expectation value of the triple product is then

$$\langle a_0 a_1 a_2 \rangle = \mathrm{Tr}\left[\rho_A \vec{a}_{012} \cdot \vec{\sigma}\right] = \vec{a}_{012} \cdot \vec{u}. \tag{29}$$

Unlike the product of two noncommuting observable in (25), the triple product in (29) depends on the Bloch vector $\vec{u}$, more in line with intuitive expectation.

Based on the commutation properties of the Pauli matrices, one can generalize these properties of the product of noncommuting observables $a_0 a_1 a_2 \ldots a_N$. If the number of operators $N+1$ is even, the symmetrization is always an operator proportional to the identity matrix, resulting in an expectation value independent of $\vec{u}$. If $N+1$ is odd, the symmetrization can be written as an operator $\vec{a}_{01\ldots N} \cdot \vec{\sigma}$, with an expectation value $\vec{a}_{01\ldots N} \cdot \vec{u}$. It is the odd case that is relevant for the Quasi-Bell inequalities we introduce.

B. Moyal characteristic function derivation

The heuristic application of symmetrization, though substantiated, may be objected to as containing an element of speculation. We therefore produce a more concrete derivation, based on Moyal's seminal representation of quantum mechanics as a statistical theory.[19] Our derivation is simpler and more flexible than that of Chandler *et al.*,[46] which applied Moyal's full Fourier approach to the limited case of mutually orthogonal spin directions.

We begin by defining the *characteristic function* of the three classical quantities a_0, a_1, a_2, given by

$$M(\theta_0, \theta_1, \theta_2) \equiv \langle e^{i(\theta_0 a_0 + \theta_1 a_1 + \theta_2 a_2)} \rangle, \tag{30}$$

where the classical expectation value is over the a_i, and the θ_i are real parameters.

The characteristic function is a standard construct in statistics,[54] useful for the calculation of moments. Of interest to us, the first triply joint moment of a_0, a_1, a_2 is given by the mixed partial derivative of the characteristic function evaluated at zero, as per

$$\langle a_0 a_1 a_2 \rangle = \left[\frac{\partial^3}{i^3 \partial\theta_0 \partial\theta_1 \partial\theta_2} M(\theta_0, \theta_1, \theta_2) \right]^{\theta_i = 0}, \tag{31}$$

where $\theta_i = 0$ denotes the evaluation $\theta_0 = \theta_1 = \theta_2 = 0$.

One can then apply Moyal's novel idea of quantizing the characteristic function to the problem at hand. We can unambiguously quantize (30) in the standard manner of replacing the classical quantities with quantum observables. In doing so, we end up with the quantum characteristic function,

$$M_\psi(\theta_0, \theta_1, \theta_2) \equiv \langle \psi | \, e^{i(\theta_0 \hat{a}_0 + \theta_1 \hat{a}_1 + \theta_2 \hat{a}_2) \cdot \vec{\sigma}} \, | \psi \rangle, \tag{32}$$

where $|\psi\rangle$ is the quantum state. Replacing M in (31) with M_ψ, we can then calculate first joint moment of the quantum characteristic function to get the *quantum expectation value of the product* as

$$\langle a_0 a_1 a_2 \rangle_\psi = \langle \psi | \left[\frac{\partial^3}{i^3 \partial\theta_0 \partial\theta_1 \partial\theta_2} e^{i(\theta_0 \hat{a}_0 + \theta_1 \hat{a}_1 + \theta_2 \hat{a}_2) \cdot \vec{\sigma}} \right]^{\theta_i = 0} | \psi \rangle. \tag{33}$$

It is clear from (33) that the quantum operator corresponding to the product $a_0 a_1 a_2$ is simply the one whose expectation value is calculated on the right hand side. Relegating the algebra to the Appendix, this operator reduces to precisely $\vec{a}_{012} \cdot \vec{\sigma}$ defined in (27). Therefore, the Moyal quantization yields exactly the same result as heuristic symmetrization. Although this equivalence is remarkable in many ways, it may be seen as following from the characteristic function's symmetry in a_0, a_1, a_2 and the subsequent application of canonical quantization.

A derivation analogous to the above can be performed for the product of any number of spin observables. The result is identical to heuristic symmetrization in each case.

We emphasize that our derivation above only made use of the existing concepts in classical probability theory and standard canonical quantization. Therefore symmetrization is not an additional concept *per se*, rather it arises from the combination of old concepts in a new way, and logically follows from established quantum theory.

That said, canonical quantization for all its successes is known to create ambiguity in some cases.[55,56] It is uncertain whether Moyal's and our choice of applying quantization to the exponential in the characteristic function are consistent with applying it to other possible functions. Therefore, although symmetrization is straightforward mathematically, this does not necessarily mean that it corresponds to physical reality.

It may be possible to test symmetrization experimentally, by evolving real systems whose Hamiltonians contain products of noncommuting observables, and comparing the experimental results with theoretical simulations based on symmetrization. If the two approaches agree, this would be the ideal justification of symmetrization. However, it is not trivial to create such Hamiltonians, making this a challenging and interesting potential research project.

VI. QUASI-BELL INEQUALITIES: THREE OBSERVABLES

We are now in a position to construct the main result of this paper, the Quasi-Bell inequalities involving products of noncommuting observables. They yield a quantum to classical violation higher than analogous Bell inequalities that do not use noncommuting products.

We start with the traditional CHSH inequality. Consider classical quantities $a_0, a_1, b_0, b_1 \in \{1, -1\}$, two each for Alice and Bob. Define $L = a_0b_0 + a_0b_1 + a_1b_0 - a_1b_1$. It can be factored to

$$L = a_0(b_0 + b_1) + a_1(b_0 - b_1). \tag{34}$$

Of the two bracketed terms in (34), one must be 0 and the other ±2. Then $L = \pm 2$ for any a_0, a_1, b_0, b_1.

Extending to the classical probabilistic case, suppose that a_0, a_1, b_0, b_1 each have some probability of being 1 or −1. In other words, there exists an ensemble of possible realizations, each element of which has $L = \pm 2$. Then the *classical expectation value* over the whole ensemble must satisfy $-2 \le \langle L \rangle \le 2$. More precisely

$$|\langle L \rangle| = |\langle a_0b_0 \rangle + \langle a_0b_1 \rangle + \langle a_1b_0 \rangle - \langle a_1b_1 \rangle| \le 2, \tag{35}$$

which is the well-known CHSH inequality.[3,57] It can be written in matrix form as

$$\left| \left\langle \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}^\dagger \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} \right\rangle \right| \le 2. \tag{36}$$

Quantizing our model, suppose that our bipartite system is made up of two spin-$\frac{1}{2}$ particles, shared between Alice and Bob, with measurement results governed by quantum theory. Let a_0, a_1 represent the spin of Alice's particle along the directions of unit vectors $\hat{a}_0, \hat{a}_1$, and b_0, b_1 represent the spin of Bob's particle along $\hat{b}_0, \hat{b}_1$. Alice (Bob) chooses the spin measurement direction $\hat{a}_0$ or $\hat{a}_1$ ($\hat{b}_0$ or $\hat{b}_1$). Assuming Alice and Bob share a singlet state $|\psi\rangle = \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)$, expectation values of joint measurements are

$$\langle (\vec{a} \cdot \vec{\sigma}) \otimes (\vec{b} \cdot \vec{\sigma}) \rangle_\psi = -\vec{a} \cdot \vec{b}, \tag{37}$$

for any $\vec{a}$ and $\vec{b}$. Choosing

$$\hat{a}_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \ \hat{a}_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \ \hat{b}_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \ \hat{b}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \tag{38}$$

and using (37) to quantize the expression (35), we have

$$|\langle L \rangle_\psi| = |-\hat{a}_0 \cdot \hat{b}_0 - \hat{a}_0 \cdot \hat{b}_1 - \hat{a}_1 \cdot \hat{b}_0 + \hat{a}_1 \cdot \hat{b}_1| = 2\sqrt{2}. \tag{39}$$

Therefore, quantum mechanics violates the classical bound of the CHSH inequality (35) by a factor of $\sqrt{2}$, its maximal possible violation,[58] famously indicating that quantum theory is not locally real.

We now introduce additional choices for Alice and Bob to measure the spin along the directions $\hat{a}_2$ and $\hat{b}_2$ with results $a_2, b_2 \in \{1, -1\}$, respectively. We call this case the second order and the CHSH inequality first order.

In (34), we made the simple yet useful observation that exactly one of the two quantities $b_0 + b_1$ and $b_0 - b_1$ was nonzero and took on the value ±2. We seek analogous quantities that include b_2. Consider the test expression

$$b_0 + b_1 + b_2 + b_0b_1b_2. \tag{40}$$

It takes the value ±4 if $b_0 = b_1 = b_2 = \pm 1$ and is 0 otherwise. Of the eight (2^3) possible realizations of the triple b_0, b_1, b_2, only two of them lead to a nonzero value for the expression (40).

We produce three additional test expressions from (40) by flipping the sign of b_1, b_2, or both. For any realization of b_0, b_1, b_2, exactly one of the four total test expressions has a nonzero value, equal to ± 4. This is outlined in Table I.

We then define the quantity L_2, as the sum of products of the four test expressions with the four factors a_0, a_1, a_2, and $a_0a_1a_2$, respectively. Since each factor has a value ± 1, and only one of the test expressions has a nonzero value ± 4, we conclude $L_2 = \pm 4$ for any given realization. Averaging all possible realizations of a_m and b_n in the ensemble, we have

$$|\langle L_2\rangle| = \left|\left\langle \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_0a_1a_2 \end{bmatrix}^\dagger \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_0b_1b_2 \end{bmatrix} \right\rangle\right| \le 4. \tag{41}$$

This matrix inequality is the sought-after generalization of (36) and must hold under assumptions of local realism. Note that the entries of the matrix in (41) are the coefficients of the individual terms in the lower part of the leftmost column in Table I.

We now quantize this expression by applying the symmetrization heuristic. Defining $\vec{a}_{012}, \vec{b}_{012}$ as in (28), and making use of (37) to evaluate the quantization of the expression in (41), we have

$$|\langle L_2\rangle_\psi| = \left|\begin{bmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \hat{a}_2 \\ \vec{a}_{012} \end{bmatrix}^\dagger \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \\ \vec{b}_{012} \end{bmatrix}\right|, \tag{42}$$

where the transpose also applies to the unit vectors inside the left supervector, and $\hat{a}_m^\dagger \hat{b}_n = \hat{a}_m \cdot \hat{b}_n$.

Next, assign the following unit vectors:

$$\hat{a}_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \hat{a}_1 = \frac{1}{2}\begin{pmatrix} 1 \\ \sqrt{3} \\ 0 \end{pmatrix}, \quad \hat{a}_2 = \frac{1}{2}\begin{pmatrix} 1 \\ -\sqrt{3} \\ 0 \end{pmatrix}, \quad \hat{b}_0 = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}, \quad \hat{b}_1 = \frac{1}{2}\begin{pmatrix} -1 \\ \sqrt{3} \\ 0 \end{pmatrix}, \quad \hat{b}_2 = \frac{1}{2}\begin{pmatrix} -1 \\ -\sqrt{3} \\ 0 \end{pmatrix}. \tag{43}$$

These are six maximally separated vectors in the plane, with an angle $\frac{\pi}{3}$ between the adjacent vectors. The inner product any two vectors is the cosine of a multiple of this angle, i.e., $\frac{1}{2}$, $-\frac{1}{2}$, or -1. Indeed $\hat{a}_0 \cdot \hat{a}_1 = \hat{a}_0 \cdot \hat{a}_2 = \hat{b}_0 \cdot \hat{b}_1 = \hat{b}_0 \cdot \hat{b}_2 = \frac{1}{2}$, $\hat{a}_1 \cdot \hat{a}_2 = \hat{b}_1 \cdot \hat{b}_2 = -\frac{1}{2}$, and the inner product matrix between the a and b vectors is

$$\begin{bmatrix} \hat{a}_0^\dagger \\ \hat{a}_1^\dagger \\ \hat{a}_2^\dagger \end{bmatrix} \begin{bmatrix} \hat{b}_0 & \hat{b}_1 & \hat{b}_2 \end{bmatrix} = \begin{bmatrix} -1 & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & -1 \\ -\frac{1}{2} & -1 & \frac{1}{2} \end{bmatrix}. \tag{44}$$

Moreover, for this assignment, (28) implies $\vec{a}_{012} = \frac{1}{6}(-\hat{a}_0 + \hat{a}_1 + \hat{a}_2) = 0$. Similarly, $\vec{b}_{012} = 0$. Therefore, evaluating (42) yields

$$|\langle L_2\rangle_\psi| = 6. \tag{45}$$

The quantized result violates the classical bound of 4 in (41) by a factor of $\frac{3}{2}$, surpassing the CHSH violation factor of $\sqrt{2}$. Quantum mechanical violation of classical bounds can be increased when

TABLE I. Values of the four test expressions for the eight possible realizations of b_0, b_1, b_2. For each realization, only one of the test expressions is nonzero and takes the value ± 4.

b_0	+1	+1	+1	+1	−1	−1	−1	−1
b_1	+1	+1	−1	−1	+1	+1	−1	−1
b_2	+1	−1	+1	−1	+1	−1	+1	−1
$b_0 + b_1 + b_2 + b_0b_1b_2$	+4	0	0	0	0	0	0	−4
$b_0 - b_1 + b_2 - b_0b_1b_2$	0	0	+4	0	0	−4	0	0
$b_0 + b_1 - b_2 - b_0b_1b_2$	0	+4	0	0	0	0	−4	0
$b_0 - b_1 - b_2 + b_0b_1b_2$	0	0	0	+4	−4	0	0	0

observables on the same subsystem (qubit) are multiplied, and their noncommutation is exploited. Put differently, noncommutativity of quantum operations contributes to the quantum violation of classical realism. Indeed, it has been argued that the product of noncommuting quantum observables in an isolated system can violate classical realism without any entanglement.[59–61]

In a similar vein to Tsirelson's bound, Lagrange multipliers show that the vectors in (43) are a local maximum of (42). Numerical optimization suggests it is a global maximum, and $|\langle L_2\rangle_\psi|$ cannot exceed 6.

Interestingly, the quantum to classical violation of the factor of $\frac{3}{2}$ is precisely the maximal quantum violation of Bell's original inequality,[2] which is realized for the same measurement directions $\hat{a}_0, \hat{a}_1$ and $\hat{a}_2$ in (43).[62,63] This is despite the very different forms of (41) and Bell's original inequality, and the presence of noncommuting products in one but not the other.

VII. QUASI-BELL INEQUALITIES: N + 1 OBSERVABLES

In search of a higher quantum to classical violation, we generalize our inequality (43) to arbitrary order. Let the Nth order inequality be constructed as follows. Suppose Alice and Bob share a bipartite system of two spin-$\frac{1}{2}$ particles in a singlet state. They each have $N+1$ measurement options, yielding results $a_0, a_1, \ldots, a_N \in \{1, -1\}$ and $b_0, b_1, \ldots, b_N \in \{1, -1\}$, respectively. The unit vectors indicating the direction of the quantized spin operator for a_m, b_n are $\hat{a}_m$, $\hat{b}_n$, respectively.

Define the scaled Hadamard matrices M and M_n as

$$M \equiv \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad M_n \equiv M^{\otimes n}, \tag{46}$$

where we used the tensor power. The matrix M_n has dimensionality $2^n \times 2^n$ and satisfies $M_n = M \otimes M_{n-1}$. For completeness, define the trivial matrix $M_0 \equiv [1]$.

Recursively define the vectors $\vec{A}_n$ and $\vec{B}_n$ as

$$\vec{A}_n \equiv \begin{bmatrix} 1 \\ a_0 a_n \end{bmatrix} \otimes \vec{A}_{n-1}, \qquad \vec{B}_n \equiv \begin{bmatrix} 1 \\ b_0 b_n \end{bmatrix} \otimes \vec{B}_{n-1}, \tag{47}$$

with $\vec{A}_0 \equiv [a_0]$, $\vec{B}_0 \equiv [b_0]$. Both $\vec{A}_n$ and $\vec{B}_n$ are of length 2^n. In the recursion, we use $a_n^2, b_n^2 = 1$ to simplify, leading to no terms having a power greater than unity. Table II lists the first few $\vec{A}_n$.

Define the Nth order quantity $K_N \equiv \vec{A}_N^\dagger M_N \vec{B}_N$. Given the above, it can be simplified as follows:

$$\begin{aligned} K_N &= \vec{A}_N^\dagger M_N \vec{B}_N \\ &= \left(\begin{bmatrix} 1 \\ a_0 a_N \end{bmatrix}^\dagger \otimes \vec{A}_{N-1}^\dagger\right) M \otimes M_{N-1} \left(\begin{bmatrix} 1 \\ b_0 b_N \end{bmatrix} \otimes \vec{B}_{N-1}\right) \\ &= a_0 b_0 \frac{1}{2}\begin{bmatrix} a_0 \\ a_N \end{bmatrix}^\dagger \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} b_0 \\ b_N \end{bmatrix} \otimes \vec{A}_{N-1}^\dagger M_{N-1} \vec{B}_{N-1} \\ &= \pm K_{N-1}, \end{aligned} \tag{48}$$

TABLE II. The vector $\vec{A}_n$ as defined in (47) for the first few n.

$\vec{A}_1$	$\vec{A}_2$	$\vec{A}_3$	$\vec{A}_4$
$\begin{bmatrix} a_0 \\ a_1 \end{bmatrix}$	$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_0a_1a_2 \end{bmatrix}$	$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_0a_1a_2 \\ a_3 \\ a_0a_1a_3 \\ a_0a_2a_3 \\ a_1a_2a_3 \end{bmatrix}$	$\begin{bmatrix} \vec{A}_3 \\ a_4 \\ a_0a_1a_4 \\ a_0a_2a_4 \\ a_1a_2a_4 \\ a_0a_3a_4 \\ a_1a_3a_4 \\ a_2a_3a_4 \\ a_0a_1a_2a_3a_4 \end{bmatrix}$

TABLE III. The maximized quantum expectation value $|\langle K_N\rangle_\psi|$ to three decimal places for N up to 10.

Order N	0	1	2	3	4	5	6	7	8	9	10
max $|\langle K_N\rangle_\psi|$	1	1.414	1.5	1.432	1.469	1.443	1.467	1.45	1.467	1.455	1.469

where in the last line we noted $a_0 b_0 = \pm 1$, and the matrix multiplication on the left side of tensor product is a CHSH term taking values ± 2. Extending the recursion,

$$K_N = \pm K_{N-1} = \cdots = \pm K_0 = \pm a_0 b_0 = \pm 1.$$

Since for each realization of a_n and b_n, K_N is ± 1, the average over the classical ensemble satisfies

$$|\langle K_N\rangle| = |\langle \vec{A}_N^\dagger M_N \vec{B}_N\rangle| \le 1. \tag{49}$$

Comparing (49) with its precursors (42) and (36), we see a power of 2 scaling difference. The factor of $\frac{1}{2}$ in the definition of M (46) cancels this power and ensures that the classical expectation value has an absolute value at most unity for all orders. This facilitates the comparison of inequalities of different orders.

Finally, we quantize the expression for K_N and maximize its quantum expectation value over all possible measurement choices $\hat{a}_n$, $\hat{b}_n$. The products of noncommuting observables are always quantized through symmetrization, analogous to (28). We obtained numerical results up to $N = 10$, shown in Table III.

The 0th order is the classical case, the 1st order violation is the CHSH value of $\sqrt{2}$, and the 2nd order yields $\frac{3}{2}$ violation demonstrated above. Interestingly, optimized violation ratios for higher orders always lie between the 1st and 2nd order cases.

The optimal measurement vectors in the 1st and 2nd order cases, in (38) and (43), respectively, were coplanar. The optimal vectors for higher orders also turn out to be coplanar. More precisely, we found that the optimal vectors satisfy $\hat{a}_2 = \hat{a}_3 = \cdots = \hat{a}_N$, $\hat{b}_2 = \hat{b}_3 = \cdots = \hat{b}_N$. Thus, for each qubit subsystem, the $N + 1$ optimized spin measurement vectors include only three unique (and coplanar) vectors, as $N - 1$ of them are identical. This gives insight into why going beyond the 2nd order actually decreases the violation ratio; adding more measurement options beyond three only replicates an existing measurement option when optimized. Further, product vectors $\vec{a}_{012}, \vec{a}_{013}, \vec{a}_{123}$, etc., cannot all simultaneously be optimized to zero as in the 2nd order case.

If this family of inequalities turns out to yield the highest possible violation, it would imply that adding a third spatial dimension to a two dimensional system does not increase the maximal possible "non-classicality." It is unclear what the effect of additional spatial dimensions beyond three would be.

VIII. LOCALITY OF WERNER STATES

Here we put the findings in perspective by comparing them to the existing work on ranges of locality. We start by adding noise to our singlet state $|\psi\rangle$, by replacing it with the two qubit Werner state,

$$\rho(z) = \frac{1-z}{4} I + z|\psi\rangle\langle\psi|.$$

This results in the expectation value (37) changing to

$$\langle(\vec{c}\cdot\vec{\sigma})\otimes(\vec{d}\cdot\vec{\sigma})\rangle_{\rho(z)} = -z\vec{c}\cdot\vec{d}. \tag{50}$$

Therefore all quantum expectation values are now scaled by a factor of z.

A rich research program has followed from investigating the lowest value of z for which the Werner state violates a Bell inequality, and the highest value for which a local hidden variable model can be constructed. With the progress of research, the two values have been converging, but have yet to meet.

The CHSH inequality implies a nonclassicality range of $z > \frac{1}{\sqrt{2}} \approx 0.7071$, Vértesi's slightly improved the range to $z > 0.7056$,[15] and Brierley *et al.* to $z > 0.7012$.[16] It is worth noting that although John Bell's original inequality matches our $\frac{3}{2}$ violation for a singlet state,[2] it loses much of its efficacy for Werner states, yielding a weaker nonclassicality range for z than the CHSH inequality.[62]

On the other end of the spectrum, the Werner state is separable, and therefore local, for $z \leq \frac{1}{3}$. By explicit construction of a simple hidden variable model, Werner showed that it is local for $z \leq \frac{1}{2}$,[21] with the surprising implication that an entangled state could have a local hidden variable description. Toner *et al.* constructed an even stronger hidden variable model demonstrating locality for $z \lesssim 0.6595$.[22,26] Most recently, Hirsch *et al.* reported locality for $z \lesssim 0.6829$,[64,65] meaning the boundary between the local and nonlocal Werner states must lie between 0.6829 and 0.7012.

It is now natural to ask where our result lies along this spectrum. The expectation value (50) changes our maximal quantum to classical violation to $\frac{3}{2}z$. This means that the Werner state violates our Quasi-Bell inequality for $z > \frac{2}{3}$. It may then seem that our result contradicts the range of locality given by Hirsch *et al.* But, in fact, this is to be expected; the additional burden of producing joint probabilities consistent with quantum theory weakens our local hidden variable models relative to the standard ones, leading to a smaller range of locality.

The above local hidden variable models by various authors do not define joint probability distributions for noncommuting observables as we do. Therefore, they should not be compared at a face value with our $z > \frac{2}{3}$ nonlocality range.

Nonetheless, addressing joint probabilities is a burden that local hidden variable models should be able to bear. It is in principle possible to extend standard local hidden variable models to do this. Defining joint probabilities (e.g., $p(a_0, a_1|\hat{a}_0, \hat{a}_1)$) that imply the marginal probabilities $(p(a_0|\hat{a}_0), p(a_1|\hat{a}_1))$ in the existing models would achieve this. The joint probabilities would be constructed to respect some desired criteria, such as positivity, the Fréchet inequalities, or, more likely, the symmetrization procedure, in which case the resulting local hidden variable model must satisfy our Quasi-Bell inequalities.

IX. DISCUSSION

Here we summarize our results, discussing their meaning and significance. We began by reviewing the concept of joint probabilities for commuting observables, and relating the quantum probabilities with those due to classical local hidden variables. We then examined joint probabilities of noncommuting observables, showing that in general they require some assumptions to determine them from the marginal probabilities.

We found that assuming the joint probability of noncommuting observables obeys the Fréchet inequalities is equivalent to assuming its positivity. It turns out that for pure single-qubit states, positivity implies independence of measurement results from noncommuting observables. Since independence is problematic, we sought other avenues to calculating the joint probability. This resulted in negative joint probabilities, well known for noncommuting observables, which do not lead to physical contradictions as they cannot be directly measured.

We introduced symmetrization as a means to quantization of classical products of noncommuting observables, and justified it heuristically then more carefully via Moyal quantization. The justification of symmetrization is interesting in its own right, independent of Bell inequalities. It may lead to its own research trajectory where one tests symmetrization through physical Hamiltonians that include products of noncommuting observables. Although symmetrization mathematically follows from the application of canonical quantization, it remains to be experimentally tested.

Based on symmetrization, we created a hierarchy of Quasi-Bell inequalities, which yields a quantum to classical violation higher than the standard Bell inequalities. Moreover, the family of Quasi-Bell inequalities in Secs. VI and VII exceeds treatments relying on Grothendieck inequalities.[66–68] This is because the latter are restricted to unit vectors, while in this work, the vectors derived from quantization of noncommuting products, such as $\vec{a}_{012}$, are not generally of unit magnitude.

Although our Quasi-Bell inequality yields a higher violation, it cannot be compared with standard Bell inequalities or local hidden variable models on Werner states, because the latter do not consider joint probabilities of noncommuting observables.

Our findings shed important light on joint distributions of noncommuting observables and their consequent extensions of Bell inequalities. The two essential "strange" features of quantum theory, noncommutativity and nonlocality, are shown to influence one another in interesting ways. The findings also help us better understand the limitations of local hidden variable models, and the possibility of extending them to include joint probabilities. We showed interesting examples of negative probabilities that appear in intermediate theoretical quantities (e.g., joint probability) but not in observable measurements.

There of course remains the important question: can Quasi-Bell inequalities be tested experimentally? Traditional Bell inequalities proved that quantum mechanics, as a theory, is nonlocal. The question then became whether reality is nonlocal, i.e., is quantum mechanics an accurate representation of reality? This led to numerous experimental tests spanning six decades[4–9,69] which confirmed the validity of quantum theory.

However, while the insights of Quasi-Bell inequalities are many, they remain, thus far, entirely on the theoretical side. Once one decides to do the experiment to measure the quantities in the Quasi-Bell inequality, one has to measure symmetrized operators like $\vec{a}_{012} \cdot \vec{\sigma}$ and $\vec{b}_{012} \cdot \vec{\sigma}$ (which are incidentally set to zero for the optimal setting for $N = 3$), because one cannot simultaneously measure noncommuting observables to calculate their product. But in deriving the symmetrization from the noncommuting product, we used canonical quantization, thereby *assuming the validity of quantum theory,* and as a consequence, nonlocality. Therefore such an experiment would be implicitly assuming nonlocality and cannot be used to test it.

This will remain true unless direct measurement of the product of noncommuting observables becomes experimentally possible. In which case direct measurement of the non commuting product will circumvent the need for symmetrization. Thankfully, this is not a serious problem; great strides by experimentalists confirming the violation of traditional Bell inequalities mean that the nonlocality of nature is no longer in doubt.

ACKNOWLEDGMENTS

We thank Toner *et al.* for sharing their unpublished manuscript.[26] We also thank the anonymous referees and Dr. Michael Hall for insightful critiques. This work was supported by the Director, Office of Science, Office of Basic Energy Sciences, of the USA Department of Energy under Contract No. DE-AC02-05CH11231 and the Division of Chemical Sciences, Geosciences and Biosciences Division, Office of Basic Energy Sciences through Grant No. DE-AC03-76F000098 (at LBNL and UC Berkeley).

APPENDIX: MOYAL QUANTIZATION OF THE PRODUCT

In this appendix, we simplify the expression in (33) to find the operator corresponding to Moyal quantization of the product $a_0 a_1 a_2$.

We define

$$\vec{\chi} \equiv \theta_0 \hat{a}_0 + \theta_1 \hat{a}_1 + \theta_2 \hat{a}_2,$$
$$\chi \equiv |\vec{\chi}| = \sqrt{\theta_0^2 + \theta_1^2 + \theta_2^2 + 2(\hat{a}_0 \cdot \hat{a}_1)\theta_0\theta_1 + 2(\hat{a}_0 \cdot \hat{a}_2)\theta_0\theta_2 + 2(\hat{a}_1 \cdot \hat{a}_2)\theta_1\theta_2}. \tag{A1}$$

Then using the well-known identity for exponents of Pauli vectors, our quantum operator in (33) becomes

$$\left[\frac{\partial^3}{i^3 \partial\theta_0 \partial\theta_1 \partial\theta_2}\left(\cos\chi\, I + i\,\frac{\sin\chi}{\chi}\,(\vec{\chi}\cdot\vec{\sigma})\right)\right]^{\theta_i=0}. \tag{A2}$$

Considering only the cosine term, we expand its Taylor series to get

$$\left[\frac{\partial^3}{i^3\partial\theta_0\partial\theta_1\partial\theta_2}\left(1-\frac{1}{2!}(\theta_0^2+\theta_1^2+\theta_2^2+2(\hat{a}_0\cdot\hat{a}_1)\theta_0\theta_1+2(\hat{a}_0\cdot\hat{a}_2)\theta_0\theta_2+2(\hat{a}_1\cdot\hat{a}_2)\theta_1\theta_2)+O(\theta^4)\right)\right]^{\theta_i=0} I, \tag{A3}$$

where $O(\theta^4)$ denotes terms with products of four or more θ_i. Since the Taylor expansion lacks a $\theta_0\theta_1\theta_2$ term, it is clear that the expression (A3) vanishes. We are then left with the more interesting sine term, which we also expand in a Taylor series and simplify as

$$\begin{aligned}
&\left[\frac{\partial^3}{i^2\partial\theta_0\partial\theta_1\partial\theta_2}\left(1-\frac{1}{3!}\left(\theta_0^2+\theta_1^2+\theta_2^2+2(\hat{a}_0\cdot\hat{a}_1)\theta_0\theta_1+2(\hat{a}_0\cdot\hat{a}_2)\theta_0\theta_2\right.\right.\right.\\
&\quad\left.\left.\left.+2(\hat{a}_1\cdot\hat{a}_2)\theta_1\theta_2)+O(\theta^4)\right)(\theta_0\hat{a}_0+\theta_1\hat{a}_1+\theta_2\hat{a}_2)\right]^{\theta_i=0}\cdot\vec{\sigma}\\
&=\frac{1}{3}\left[\frac{\partial^3}{\partial\theta_0\partial\theta_1\partial\theta_2}((\hat{a}_0\cdot\hat{a}_1)\theta_0\theta_1+(\hat{a}_0\cdot\hat{a}_2)\theta_0\theta_2+(\hat{a}_1\cdot\hat{a}_2)\theta_1\theta_2)(\theta_0\hat{a}_0+\theta_1\hat{a}_1+\theta_2\hat{a}_2)\right]^{\theta_i=0}\cdot\vec{\sigma}\\
&=\frac{1}{3}\left[(\hat{a}_1\cdot\hat{a}_2)\hat{a}_0+(\hat{a}_2\cdot\hat{a}_0)\hat{a}_1+(\hat{a}_0\cdot\hat{a}_1)\hat{a}_2\right]\cdot\vec{\sigma}\\
&=\vec{a}_{012}\cdot\vec{\sigma},
\end{aligned} \tag{A4}$$

where in going to the second line we noted that θ_i^2 and $O(\theta^4)$ terms cannot contribute to the $\theta_0\theta_1\theta_2$ term, and the final line is exactly the one that is defined in (27).

[1] A. Einstein, B. Podolsky, and N. Rosen, Phys. Rev. **47**, 777 (1935).
[2] J. S. Bell, Physics **1**, 195 (1964).
[3] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).
[4] S. J. Freedman and J. F. Clauser, Phys. Rev. Lett. **28**, 938 (1972).
[5] A. Aspect, P. Grangier, and G. Roger, Phys. Rev. Lett. **47**, 460 (1981).
[6] A. Aspect, P. Grangier, and G. Roger, Phys. Rev. Lett. **49**, 91 (1982).
[7] A. Aspect, J. Dalibard, and G. Roger, Phys. Rev. Lett. **49**, 1804 (1982).
[8] B. Hensen, H. Bernien, A. E. Dréau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenberg, R. F. L. Vermeulen, R. N. Schouten, C. Abellán, W. Amaya, V. Pruneri, M. W. Mitchell, M. Markham, D. J. Twitchen, D. Elkouss, S. Wehner, T. H. Taminiau, and R. Hanson, Nature **526**, 682 (2015).
[9] L. K. Shalm, E. Meyer-Scott, B. G. Christensen, P. Bierhorst, M. A. Wayne, M. J. Stevens, T. Gerrits, S. Glancy, D. R. Hamel, M. S. Allman, K. J. Coakley, S. D. Dyer, C. Hodge, A. E. Lita, V. B. Verma, C. Lambrocco, E. Tortorici, A. L. Migdall, Y. Zhang, D. R. Kumor, W. H. Farr, F. Marsili, M. D. Shaw, J. A. Stern, C. Abellán, W. Amaya, V. Pruneri, T. Jennewein, M. W. Mitchell, P. G. Kwiat, J. C. Bienfang, R. P. Mirin, E. Knill, and S. W. Nam, Phys. Rev. Lett. **115**, 250402 (2015).
[10] S. L. Braunstein and C. M. Caves, Ann. Phys. **202**, 22 (1990).
[11] D. Avis, H. Imai, T. Ito, and Y. Sasaki, J. Phys. A: Math. Gen. **38**, 10971 (2005).
[12] N. Gisin, in *Quantum Reality, Relativistic Causality, and Closing the Epistemic Circle* (Springer, Netherlands, 2009), pp. 125–138.
[13] D. Collins, N. Gisin, N. Linden, S. Massar, and S. Popescu, Phys. Rev. Lett. **88**, 040404 (2002).
[14] M. Junge, C. Palazuelos, D. Perez-Garcia, I. Villanueva, and M. M. Wolf, Commun. Math. Phys. **300**, 715 (2010).
[15] T. Vértesi, Phys. Rev. A **78**, 032112 (2008).
[16] S. Brierley, M. Navascues, and T. Vertesi, "Convex separation from convex optimization for large-scale problems," e-print arXiv:1609.05011.
[17] F. Arnault, J. Phys. A: Math. Theor. **45**, 255304 (2012).
[18] A. Fine, J. Math. Phys. **23**, 1306 (1982).
[19] J. Moyal, Math. Proc. Cambridge Philos. Soc. **45**, 99 (1949).
[20] A. O. Barut, M. Božić, and Z. Marić, Found. Phys. **18**, 999 (1988).
[21] R. F. Werner, Phys. Rev. A **40**, 4277 (1989).
[22] A. Acín, N. Gisin, and B. Toner, Phys. Rev. A **73**, 062105 (2006).
[23] R. Augusiak, M. Demianowicz, and A. Acín, J. Phys. A: Math. Theor. **47**, 424002 (2014).
[24] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Rev. Mod. Phys. **86**, 419 (2014).
[25] O. Gamel, Phys. Rev. A **93**, 062320 (2016).
[26] B. Toner, A. Acín, and N. Gisin, "Grothendieck Constants and Local Models for Werner and other noisy quantum states" (unpublished).
[27] C. Y. She and H. Heffner, Phys. Rev. **152**, 1103 (1966).
[28] E. Prugovecki, Found. Phys. **3**, 3 (1973).
[29] L. E. Ballentine, Rev. Mod. Phys. **42**, 358 (1970).
[30] M. Fréchet, Fundamenta Mathematicae **25**, 379 (1935).
[31] M. Fréchet, Ann. Univ. Lyon Sect. A **14**(3), 53 (1951).
[32] S. W. Al-Safi and A. J. Short, Phys. Rev. Lett. **111**, 170403 (2013).

[33] R. P. Feynman, in *Quantum Implications: Essays in Honour of David Bohm*, edited by B. Hiley and F. D. Peat (Routledge & Kegan Paul Ltd., London, New York, 1987), pp. 235–248.
[34] L. de Broglie, *Physics and Microphysics* (Harpers, 1960).
[35] H. Margenau and R. N. Hill, Prog. Theor. Phys. **26**, 722 (1961).
[36] A. Khrennikov, Can. J. Phys. **75**, 291 (1997).
[37] A. C. Levy, "Negative probabilities in physics: A review," Ph.D. thesis, Imperial College London, 2015.
[38] K. Wódkiewicz, Phys. Lett. A **129**, 1 (1988).
[39] E. C. G. Sudarshan and T. Rothman, Int. J. Theor. Phys. **32**, 1077 (1993).
[40] T. Rothman and E. C. G. Sudarshan, Int. J. Theor. Phys. **40**, 1525 (2001).
[41] R. W. Spekkens, Phys. Rev. Lett. **101**, 020401 (2008).
[42] V. Veitch, C. Ferrie, D. Gross, and J. Emerson, New J. Phys. **14**, 113011 (2012).
[43] E. Wigner, Phys. Rev. **40**, 749 (1932).
[44] R. L. Hudson, Rep. Math. Phys. **6**, 249 (1974).
[45] J. S. Bell, Rev. Mod. Phys. **38**, 447 (1966).
[46] C. Chandler, L. Cohen, C. Lee, M. Scully, and K. Wódkiewicz, Found. Phys. **22**, 867 (1992).
[47] M. O. Scully and K. Wódkiewicz, Found. Phys. **24**, 85 (1994).
[48] M. O. Terra Cunha, V. I. Man'ko, and M. O. Scully, Found. Phys. Lett. **14**, 103 (2001).
[49] J. R. Shewell, Am. J. Phys. **27**, 16 (1959).
[50] H. Margenau and L. Cohen, in *Studies in Foundations, Methodology and Philosophy of Science* (Springer-Verlag, Berlin, 1967), Vol. 2, pp. 71–89.
[51] S. J. Gustafson and I. M. Sigal, *Mathematical Concepts of Quantum Mechanics* (Springer-Verlag, Berlin Heidelberg, 2011).
[52] J.-M. Schwindt, *Conceptual Basis of Quantum Mechanics* (Springer, Switzerland, 2016).
[53] J. von Neumann, *Mathematische Grundlagen der Quantenmechanik* (Julius Springer-Verlag, Berlin, 1932).
[54] E. Lukacs, Adv. Appl. Probab. **38**, 1 (1972).
[55] H. Groenewold, Physica **12**, 405 (1946).
[56] M. J. Gotay, H. Grundling, and C. A. Hurst, Trans. Am. Math. Soc. **348**, 1579 (1996).
[57] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, 2nd ed. (Cambridge University Press, Cambridge, 2011).
[58] B. S. Cirel'son, Lett. Math. Phys. **4**, 93 (1980).
[59] A. Fine, Phys. Rev. Lett. **48**, 291 (1982).
[60] J. D. Malley, Phys. Rev. A **69**, 022118 (2004).
[61] J. D. Malley and A. Fine, Phys. Lett. A **347**, 51 (2005).
[62] I. Pitowsky, J. Math. Phys. **49**, 012101 (2008).
[63] R. A. Bertlmann, J. Phys. A: Math. Theor. **47**, 424007 (2014).
[64] F. Hirsch, M. T. Quintino, T. Vértesi, M. Navascués, and N. Brunner, Quantum **1**, 3 (2017).
[65] F. Hirsch, M. T. Quintino, T. Vértesi, M. F. Pusey, and N. Brunner, Phys. Rev. Lett. **117**, 190402 (2016).
[66] A. Grothendieck, Bol. Soc. Mat. São Paulo **8**, 1 (1953).
[67] J. Krivine, Adv. Math. **31**, 16 (1979).
[68] B. S. Tsirel'son, J. Sov. Math. **36**, 557 (1987).
[69] M. Giustina, M. A. M. Versteegh, S. Wengerowsky, J. Handsteiner, A. Hochrainer, K. Phelan, F. Steinlechner, J. Kofler, J.-Å. Larsson, C. Abellán, W. Amaya, V. Pruneri, M. W. Mitchell, J. Beyer, T. Gerrits, A. E. Lita, L. K. Shalm, S. W. Nam, T. Scheidl, R. Ursin, B. Wittmann, and A. Zeilinger, Phys. Rev. Lett. **115**, 250401 (2015).