PHYSICAL REVIEW A **92**, 030103(R) (2015)

Limited preparation contextuality in quantum theory and its relation to the Cirel'son bound

Manik Banik, Some Sankar Bhattacharya, Amit Mukherjee, and Arup Roy
Physics and Applied Mathematics Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata 700108, India

Andris Ambainis and Ashutosh Rai
Faculty of Computing, University of Latvia, Raina bulv. 19, Riga, LV-1586, Latvia

(Received 29 June 2015; published 24 September 2015)

The Kochen–Specker (KS) theorem lies at the heart of the foundations of quantum mechanics. It establishes the impossibility of explaining predictions of quantum theory by any noncontextual ontological model. Spekkens generalized the notion of KS contextuality in [Phys. Rev. A **71**, 052108 (2005)] for arbitrary experimental procedures (preparation, measurement, and transformation procedures). Interestingly, later on it was shown that preparation contextuality powers parity-oblivious multiplexing [Phys. Rev. Lett. **102**, 010401 (2009)], a two-party information theoretic game. Thus, using resources of a given operational theory, the maximum success probability achievable in such a game suffices as a *bona fide* measure of preparation contextuality for the underlying theory. In this work we show that preparation contextuality in quantum theory is more restricted compared to a general operational theory known as *box world*. Moreover, we find that this limitation of quantum theory implies the quantitative bound on quantum nonlocality as depicted by the Cirel'son bound.

DOI: 10.1103/PhysRevA.92.030103 PACS number(s): 03.65.Ud, 03.67.Ac, 03.67.Mn

Quantum mechanics (QM) departs fundamentally from the well-known *local-realistic* world view of classical physics. This stark contrast of quantum theory with classical physics was illuminated by Bell [1]. Since the Bell's seminal work, nonlocality remains at the center of quantum foundational research [2,3]. More recently quantum nonlocality has been also established as a key resource for device-independent information technology [3,4]. Quantum nonlocality does not contradict the relativistic causality principle; however, QM is not the only possible theory that exhibits nonlocality along with satisfying the no-signaling principle; there can be nonquantum no-signaling correlations exhibiting nonlocality. One extreme example of such a correlation (more nonlocal than QM) was first constructed by Popescu and Rohrlich (PR) [5]. Whereas the PR correlation violates the Bell–Clauser–Horne–Shimony–Holt (Bell-CHSH) [1,6] inequality by algebraic maximum, the optimal Bell-CHSH violation in quantum theory is restricted by the Cirel'son bound [7]. In this work, we show that Cirel'son limit on nonlocal behavior of quantum theory can be explained from another of its very interesting features; namely, restricted *preparation contextuality*.

Nearly at the same time of Bell's result, Kochen and Specker proved another important no-go theorem showing that predictions of *sharp* (projective) measurements in QM cannot be reproduced by any noncontextual ontological model [8]. Unlike Bell-nonlocality, the structure of QM is implicit in the definition of KS contextuality. However, recently the idea of KS contextuality has been generalized by Spekkens [9] for arbitrary operational theories rather than just quantum theory and for arbitrary experimental procedures rather than just sharp measurements. It was then shown that mixed preparations (density matrices) in quantum theory exhibit preparation contextuality [9,10]. Interestingly, invoking another nonclassical concept called *steering* [11,12] along with this new idea of preparation contextuality one can establish nonlocality of QM without using any Bell-type inequalities; it has been shown that nonlocality of some hidden-variable models, underlying QM, directly follows from the steerability of bipartite pure entangled states and the preparation contextuality of mixed states [10,13,14].

The traditional definition of contextuality addresses only the contexts of projective measurements, which have been studied in much depth [15,16]. However, the generalized notion of contextuality developed by Spekkens define three different types of contexts: measurement (generalized), preparation, and transformation contexts [9]. Interest in studying contextuality in this general framework is relatively new and growing [10,13,14,17–20]. This generalized approach has led to designing more robust experimental tests of contextuality [18–20]; these recent results are very promising given that previous requirements for testing contextuality in experiments has been a topic of much controversy (for more discussion, see Ref. [20], and relevant references therein).

Our work here is build upon the notion of preparation contextuality, which addresses the impossibility of representing two equivalent preparation procedures, of an operational theory equivalently in some ontological model. More precisely, suppose we have two equivalent operational preparations P, P', i.e., the outcome probabilities $p(k|P,T,M) = p(k|P',T,M)$ $\forall$ outcomes k, transformations T, and measurements M. Then a hidden-variable (ontic) model, which reproduces $p(k|P,T,M)$ by averaging over the ontic states λ is preparation noncontextual if $\forall\ M,T : p(k|P,T,M) = p(k|P',T,M) \Rightarrow p(\lambda|P) = p(\lambda|P')$, where $p(\lambda|P)$ and $p(\lambda|P')$ represent respective distributions over the ontic states followed by operational preparations P and P' [9].

Preparation contextuality has operational usefulness because it powers *parity-oblivious multiplexing* (POM), a two-party secure computation task [17]; in this work, Spekkens *et al.* derived a "noncontextuality inequality" which places an upper bound on any operational theory that admits a preparation noncontextual ontological model. Furthermore, the authors showed that the success rate of the POM game played with only classical resources is restricted by the same inequality. Thus, Spekkens and coauthors concluded that any operational theory is preparation contextual if it can beat the classical bound in a POM task.

It turns out that, in performing a POM task, certain quantum resources can do better than any classical resource, thus proving that QM is a preparation-contextual theory. Interestingly, in this work we show that, for performing a POM task, there exists an operational (toy) theory; namely, *box world* [21,22], which can do better than QM. Thus, although QM is preparation contextual, the amount of preparation contextuality in QM is constrained compared to the box world. Furthermore, we show that the restricted preparation contextuality of quantum theory leads to its limited nonlocal behavior as depicted in the Cirel'son bound. Therefore, our result brings the qualitative connection between preparation contextuality and nonlocality explored in Refs. [10,13,14] to a quantitative footing.

Parity-oblivious multiplexing. Parity-oblivious multiplexing is a variant of the well-studied information-processing task called *random access code* (AC) [23–25]. Suppose an n-bit string x, chosen uniformly at random from $\{0,1\}^n$, is given to Alice. An integer y, chosen uniformly at random from $\{1,2,\dots,n\}$, is given to Bob, whose task now task is to guess the yth bit of Alice's input. Let us denote Bob's guess as β_y. In the POM game Alice and Bob collaborate to optimize the guessing probability $p(\beta_y = y\text{th bit of Alice})$. Alice can send to Bob any information which encodes her input. However, there is a cryptographic constraint: no information about any parity of x can be transmitted to Bob. More specifically, letting $s \in \text{Par}$ where $\text{Par} \equiv \{r | r \in \{0,1\}^n,\ \sum_i r_i \geqslant 2\}$ is the set of n-bit strings with at least 2 bits that are 1, no information about $x \cdot s = \oplus_i x_i s_i$ (termed the s parity) for any such s can be transmitted to Bob (here $\oplus$ denotes sum modulo 2).

The main result of Spekkens *et al.* [17] can be now stated more precisely: *for an n-bit POM game played with states (resources) from a preparation-noncontextual theory, the average success probability is bounded as follows:*

$$p_{NC}\big(\beta_y = y\text{th bit of Alice}\big) \leqslant \frac{1}{2}\left(1 + \frac{1}{n}\right).$$

Motivated by this result, in our work, we define the maximum success probability in a POM task in an operational theory as a *bona fide* measure to quantify the strength of preparation contextuality of the concerned theory. The approach we adopt here is similar to defining the strength of nonlocality of correlations as the amount of Bell-CHSH violation (or maximum success probability in a Bell-CHSH game). In the remainder of this paper, we focus on a 2-bit POM task. For the 2-bit POM game we adopt two different schemes: (1) an encoding-decoding scheme and (2) a correlation-assisted scheme.

(1) Encoding-decoding scheme. Alice and Bob can perform a POM task by using resources of a general operational theory. In an operational theory, the primitives of description are preparations and measurements (for simplicity, here we do not consider dynamics or transformations of the system)[21,22,26–30]. The theory simply provides an algorithm for calculating the probability $p(k|P,M)$ of an outcome k of measurement M given a preparation (state) P. The collection of all states in which the system can be prepared forms a compact and convex subset Ω of a finite-dimensional vector space V. Results of a measurement on any state ω of the theory is described by an *effect* $e : \Omega \to [0,1]$, which is a map such that $e(\omega)$ is the probability of obtaining the outcome e. There is a *unit effect* u such that $u(\omega) = 1\ \forall\ \omega \in \Omega$. Any measurement can now be expressed as some set of effects $\{e_i\}$ such that $\sum_i e_i = u$.

Alice, depending on the input string $x \in \{00,01,10,11\}$ that is given to her uniformly at random, implements a preparation procedure $P_x \in \Omega_A$ in an operational theory $\mathcal{T}$ and sends the encoded particle to Bob. For each integer $y \in \{1,2\}$, Bob implements a binary-outcome measurement M_y and reports the outcome as his output. The *average* probability of winning is given by

$$\begin{aligned} p_{\mathcal{T}}(\beta_y &= y\text{th bit of Alice}) \\ &\equiv p(\beta_y = x_y) \\ &= \frac{1}{8}\sum_{y=1}^{2}\sum_{x\in\{00,01,10,11\}} p(\beta_y = x_y|P_x,M_y). \end{aligned} \tag{1}$$

The optimal success probability in an operational theory is $p^{\text{opt}}_{\mathcal{T}}(\beta_y = x_y) := \max_{P_x,M_y} p(\beta_y = x_y)$, where optimization is performed over all possible encoding and decoding procedures allowed in the theory $\mathcal{T}$. Of course the encoding and decoding must satisfy the parity-oblivious constraint, expressed here as

$$\begin{aligned} p(P_{00}|k,M) &+ p(P_{11}|k,M) \\ &= p(P_{01}|k,M) \\ &\quad + p(P_{10}|k,M)\ \forall\ M \in \mathcal{M} \text{ and } \forall\ k. \end{aligned} \tag{2}$$

First we show that the optimal success probability in the box world is strictly greater than that of quantum theory, i.e., $p^{\text{opt}}_{\text{box}}(\beta_y = x_y) > p^{\text{opt}}_{Q}(\beta_y = x_y)$. To prove this result we first consider the quantum case, and then the box world.

(a) Quantum theory. Alice encodes her two bits into the four pure qubits with Bloch vectors $\{(\pm 1,0,0),(0,0,\pm 1)\}$ equally distributed on the equatorial X-Z plane of the Bloch sphere; as shown in Fig. 1. Bob performs the measurement $(\sigma_x + \sigma_z)/\sqrt{2}$ if he wishes to learn the first bit, and the measurement $(\sigma_x - \sigma_z)/\sqrt{2}$ if he wishes to learn the second. He guesses the bit value 0 upon obtaining the positive outcome, otherwise he guesses the bit value 1. In all cases, the guessed value is correct with probability $\frac{1}{2}(1 + \frac{1}{\sqrt{2}})$, which results in

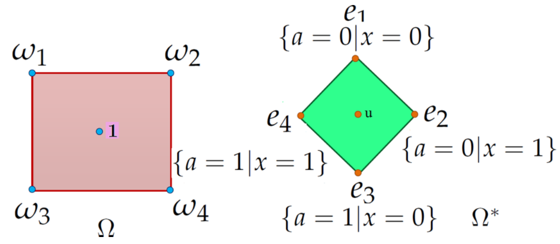

FIG. 1. (Color online) Ω denotes the normalized state space for the box world. The four corners denote four deterministic states and the central dot denotes the completely mixed state. Ω^* denotes the space of effects for the box world. $\{e_i | i = 1,\dots,4\}$ are four extremal effects of the box world. Two ideal measurements are $e_1 + e_3 = \mathbf{u} = e_2 + e_4$.

the average success probability $p_Q(\beta_y = x_y) = \frac{1}{2}(1+\frac{1}{\sqrt{2}}) > \frac{2}{3} = p_{NC}^{\text{opt}}(\beta_y = x_y)$. Since the parity 0 and parity 1 mixtures in this protocol are represented by the same density operator, no information about the parity can be obtained by any quantum measurement. Interestingly, the qubit protocol just described turns out to be quantum optimal.

Proposition 1. In a 2-bit POM game, optimum average success probability over all quantum encoding-decoding schemes is $p_Q^{\text{opt}}(\beta_y = x_y) = \frac{1}{2}(1+\frac{1}{\sqrt{2}})$.

Proof. Alice prepares and encodes as $\{ij \longmapsto \rho_{ij} : i,j \in \{0,1\}\}$, where ρ_{ij} are state operators acting on $\mathbb{C}^d$; she can always find an appropriate pure state $|\Psi_{12}\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$ and projectors $P_A, P_{A'}$ such that $\frac{1}{2}\rho_{00} = \text{tr}_1\{(P_A \otimes I)|\Psi_{12}\rangle\}$, $\frac{1}{2}\rho_{11} = \text{tr}_1\{[(I-P_A)\otimes I]|\Psi_{12}\rangle\}$, $\frac{1}{2}\rho_{01} = \text{tr}_1\{(P_{A'} \otimes I)|\Psi_{12}\rangle\}$, $\frac{1}{2}\rho_{10} = \text{tr}_1\{[(I-P_{A'})\otimes I]|\Psi_{12}\rangle\}$. Alice performs one of the following projective measurements: (i) $P_A \otimes I + (I-P_A)\otimes I = I \otimes I$, (ii) $P_{A'} \otimes I + (I-P_{A'})\otimes I = I\otimes I$, on part one of $|\Psi_{12}\rangle$ and depending on the measurement result she sends part two to Bob (or discards it). Measurements (i) and (ii) respectively produce two decompositions $\frac{1}{2}\rho_{00} + \frac{1}{2}\rho_{11}$ and $\frac{1}{2}\rho_{01} + \frac{1}{2}\rho_{10}$ for part two of $|\Psi_{12}\rangle$. Alice in this way prepares and sends ρ_{ij} to Bob, assured that the parity-oblivious condition is always satisfied.

Bob, on receiving part two, performs a two-outcome projective measurement $\{P_B,(I-P_B)\}$ $[\{P_{B'},(I-P_{B'})\}]$, if he is asked to guess Alice's first (second) bit, and answers 0 (1) when measurement outcome is $+1$ (-1). Here, P_A, $P_{A'}$, P_B, $P_{B'}$ are projectors acting on $\mathbb{C}^d$. Due to the lack of constraint on the dimension of the Hilbert space, Neumark's theorem allows us to consider only projective measurements, without loss of generality. Substituting ρ_{ij} in terms of $|\Psi_{12}\rangle$, P_A and $P_{A'}$ in the expression for average success probability for the 2-bit POM game we get $p_Q = \frac{1}{8}[4 + \langle\Psi_{12}|\{A\otimes B + A'\otimes B + A\otimes B' - A'\otimes B'\}|\Psi_{12}\rangle]$, where $A = 2P_A - I$, $A' = 2P_{A'} - I$, $B = 2P_B - I$, $B' = 2P_{B'} - I$ (see Ref. [31] for details). Since all four operators (observables) $\{A,A',B,B'\}$ have eigenvalues $\{\pm 1\} \in [-1,1]$, and any operator from the set $\{A,A'\}$ commutes with any operator from the set $\{B,B'\}$, by applying Cirel'son's result [7] it follows that $\langle\Psi_{12}|\{A\otimes B + A'\otimes B + A\otimes B' - A'\otimes B'\}|\Psi_{12}\rangle \leqslant 2\sqrt{2}$. This gives, $p_Q \leqslant \frac{1}{2}[1+\frac{1}{\sqrt{2}}]$. We have already discussed that there exists a quantum protocol to achieve this upper bound. ■

(b) Box world. Interestingly, one can exceed the optimal quantum bound in the box world. This system can be understood as a black box taking a binary input $x = 0,1$ and returning a binary output $a = 0,1$ [22]. The state of the system is described by a conditional probability distribution $P(a|x)$. The normalized state space Ω of the system can be represented as a square in $\mathbf{R}^2$ (see Fig. 1). The system thus features four pure states $\{\omega_j | j = 1,\ldots,4\}$. For each pure state, the outcome "a" is a deterministic function of the input "x" ($\omega_1 \to a = 0$, $\omega_2 \to a = x$, $\omega_3 \to a = 1$, and $\omega_4 \to a = x \oplus 1$). The center of Ω is the maximally mixed state; that is, where a is independent of x and random. This maximally mixed state has a nonunique decomposition in term of pure states, i.e., $\frac{1}{2}\omega_1 + \frac{1}{2}\omega_3 = \frac{1}{2}\omega_2 + \frac{1}{2}\omega_4 = \mathbf{1}$. The space of effects, Ω^*, is the dual of Ω (see Fig. 1). It features four extremal effects $\{e_j | j = 1,\ldots,4\}$ which correspond to the four measurement outcomes, i.e., obtaining output a for a given input x. The probability of e_j on any state is easily determined. For instance, effect e_1 has a probability one for states ω_1, ω_2 and probability zero for states ω_3, ω_4. There are two pure measurements for this system: the first is composed of effects e_1 and e_3 and corresponds to input $x = 0$; the second is composed of effects e_2 and e_4 and corresponds to input $x = 1$. Note that $e_1 + e_3 = e_2 + e_4 = \mathbf{u}$, where $\mathbf{u}$ is the unit effect. This system is also known as a generalized bit (g-bit) [21].

For performing the 2-bit POM with better-than-quantum success, Alice and Bob pursue the following strategy in the box world: Alice encodes her strings as

$$\begin{aligned} 00 &\to \omega_1, \quad 11 \to \omega_3, \\ 01 &\to \omega_2, \quad 10 \to \omega_4, \end{aligned} \tag{3}$$

and sends the encoded g-bit to Bob. The parity-oblivious condition is satisfied because $\frac{1}{2}\omega_1 + \frac{1}{2}\omega_3 = \frac{1}{2}\omega_2 + \frac{1}{2}\omega_4 = \mathbf{1}$. To decode Alice's message, Bob performs (i) measurement $\{e_1,e_3|e_1+e_3 = u\}$ if he wishes to learn the first bit and (ii) measurement $\{e_2,e_4|e_2+e_4 = u\}$ if he wishes to learn the second bit and he guesses Alice's bit as the measurement result. In every case, the guessed value is correct with certainty resulting from $p_{\text{box}}(\beta_y = x_y) = 1 > \frac{1}{2}(1+\frac{1}{\sqrt{2}}) = p_Q^{\text{opt}}(\beta_y = x_y)$. Clearly $p_{\text{box}}^{\text{opt}}(\beta_y = x_y) = 1$ as the said strategy achieves 100% success probability.

(2) Correlation-assisted scheme. Let Alice and Bob now follow a different scheme in which, instead of sending encoded states, they use correlations of preshared bipartite states allowed in an operational theory. In an operational theory a general bipartite correlation can be thought as a probability distribution $p(\mathcal{C},\mathcal{D}|\mathcal{U},\mathcal{V}) \equiv \{p(c,d|u,v)\}$, where $u \in \mathcal{U}$, $v \in \mathcal{V}$ are inputs given to Alice and Bob, respectively, and $c \in \mathcal{C}, d \in \mathcal{D}$ denote their respective outcomes. *No-signaling* correlations satisfy the conditions $p(c|u) = \sum_{d\in\mathcal{D}} p(c,d|u,v)\ \forall\, c,u,v$ and vice versa. If local outcomes depend only on the choice of local measurements and (possibly) on the value of some shared (hidden) variable $\lambda \in \Lambda$ which takes values according to some distribution $p(\Lambda) = \{p(\lambda)\}$ then the correlation is called *local*, i.e., $p_L(c,d|u,v) = \sum_{\lambda\in\Lambda} p(\lambda)p(c|u,\lambda)p(d|v,\lambda)$ for all c, d, u, v. Correlations which cannot be expressed in such form are called *nonlocal* [1]. Entangled quantum particles [32] exhibit nonlocal properties whereas they satisfy the no-signaling conditions.

By using bipartite correlations Alice and Bob can perform the POM task in the following manner: Alice, prior to starting the POM game, shares with Bob a correlated pair of particles prepared in the state (preparation) $P_{AB} \in \Omega_{AB}$. Depending on the input string given to her, Alice performs a measurement on her particle of the correlated pair and sends the measurement result to Bob via classical communications (CCs). Bob, receiving the CC from Alice, performs operations on his particle and tries to guess Alice's bit. However, the CC should not contain any information about the parity of Alice's input string. It turns out that local correlations are not useful for performing the POM task.

Proposition 2. Correlations having a local description when assisted with classical communications are not useful for performing the parity-oblivious multiplexing task.

Proof. Any local correlation between Alice and Bob can be thought of as a shared random variable $\lambda \in \Lambda$ taking values according to a probability distribution $p(\Lambda)$. Reference [17] shows that the only classical encodings of x that reveal no information about any parity (while encoding some information about x) are those that encode only a single bit x_i for some i. For simplicity, without loss of generality, consider that the shared variable takes discrete values λ_i with $\sum_{i=1}^{n} p(\lambda_i) = 1$. If the variable takes the value λ_k then Alice encodes her kth bit and sends it to Bob. Bob, if asked, can correctly reveals this kth bit while for him other bits ($i \neq k$) are completely random. Thus, by using local correlation (shared randomness), Alice and Bob can design a strategy for determining only one bit with certainty. But, whichever bit Bob guesses correctly, the (average) success probability of Bob's guess is bounded by the noncontextual bound $\frac{1}{2}(1+\frac{1}{n})$. Therefore, due to convexity of the distribution $p(\Lambda)$, it follows that any local correlation cannot beat the noncontextual bound. ■

Remark. Here it is important to note that, in a POM game, to obtain greater success than the classical bound, the theory need not contain nonlocal correlations. For example, consider a theory in which individual state space is identical as for quantum state space but the composite state space is severely more restricted than quantum state space. The state space of the composite system is a *minimal tensor product* [27] of individual Hilbert spaces and hence contains only separable states and hence no nonlocal correlation. In such a theory one can obtain the success probability of the POM game as much as quantum theory by following the optimal encoding-decoding scheme of quantum theory. What Proposition 2 proves is that, if one wants to play the POM game by using the correlation of such a local theory, she or he will not get any advantage.

Proposition 3. Any no-signaling correlation $\{p(ab|xy): a,b,x,y \in \{0,1\}\}$ violating the Bell-CHSH inequality can exceed the classical bound for performing the 2-bit parity-oblivious multiplexing task. Moreover, if, by using a correlation, the average success in 2-bit POM game exceeds the quantum limit, then nonlocality of such correlation must exceed the Cirel'son bound.

Outline of proof. A proof follows from (i) using the protocol for 2-bit random access code scenario discussed by Pawlowaski *et al.* [33] in the context of information causality, and (ii) showing that this protocol respects the parity-obliviousness condition. Then this implies that any nonlocal correlation can achieve more than the classical limit for a 2-bit POM task. Moreover, the quantum limit for a 2-bit POM game nonlocality of correlation is restricted by the Cirel'son bound. We give a complete proof of the proposition in the supplementary material [34]. ■

In the quantum world, by using correlations of entangled particles, Alice and Bob can win the POM game with more-than-classical (noncontextual) success probability. By using the steerability [11,12] of the entangled state and classical communication, Alice tries to prepare Bob's state in different preparations depending on the input string given to her. For achieving the best result, Alice attempts to prepare Bob's particle into states which achieve optimal success probability in the encoding-decoding scheme. By sharing the two-qubit maximally entangled state, Alice can prepare the optimal states by echoing an identical procedure as in the remote-state-preparation protocol [35] (see Ref. [36] for the protocol).

However, the presence of steering alone in a theory is not sufficient for achieving more than classical success probability—the theory must also be preparation contextual. For instance, there exists hypothetical *toy bit* theory [37] which allows steering, but the success probability of POM in this theory is restricted to the classical bound because the theory is preparation noncontextual (see Ref. [38]). On the other hand, although steerability in quantum theory is maximal, the optimal success probability of the POM task is restricted due to its limited preparation contextuality.

To conclude, in this work we considered an operational way to quantify the preparation contextuality of a general theory. We show that quantum theory turns out to be less preparation contextual than another operational theory; namely, the box world. Furthermore, we show that, in the quantum world, the restricted Bell-CHSH violation follows from the limited preparation contextuality of the theory. Many researchers have tried to explain the limits of the nonlocal feature in QM starting from a number of physically motivated ideas or principles. In particular, by considering various approaches, it has been successfully explained why the Bell-CHSH quantity in quantum theory is restricted to the Cirel'son bound [33]. Having established a link between the concept of nonlocality and preparation contextuality it would be interesting to suggest physical principal(s) leading to quantum bound on preparation contextuality.

Recently, the authors of Ref. [39] have shown that even-parity-oblivious encodings are equivalent to the INDEX game, which implies the $2 \to 1$ POM game is equivalent to the well-known Bell-CHSH nonlocal game. Therefore, a quantum encoding of $2 \to 1$ POM with average success probability p_Q exists only if a quantum strategy for playing the Bell-CHSH game with the same average success probability exists. We take a different approach: by maximizing over all possible encoding-decoding schemes allowed in QM, we find the optimal success probability of the 2-bit POM game: it turns out to be restricted compared to a more general operational theory. We conclude that restricted preparation contextuality, therefore, bounds the winning probability of Bell-CHSH game (nonlocality) in quantum theory.

Thanks to Guruprasad Kar for many stimulating discussions. M.B. would like to acknowledge useful conversations with Rajjak Gazi. A.M. thanks the Council of Scientific and Industrial Research, India for financial support through a Senior Research Fellowship (Grant No. 09/093(0148)/2012-EMR-I). As.R. would like to thank Andreas Winter for helping to prove the quantum optimality of 2-bit POM games. A.A. and As.R. acknowledge support by the European Union Seventh Framework Programme (FP7-ICT) under the RAQUEL (Grant Agreement No. 323970) project, QALGO (Grant Agreement No. 600700) project, and the ERC Advanced Grant MQC.

[1] J. S. Bell, Physics **1**, 195 (1964).
[2] N. D. Mermin, Rev. Mod. Phys. **65**, 803 (1993).
[3] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Rev. Mod. Phys. **86**, 419 (2014).
[4] V. Scarani, Acta Phys. Slovaca **62**, 347 (2012).
[5] S. Popescu and D. Rohrlich, Found. Phys. **24**, 379 (1994).
[6] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).
[7] B. S. Cirel'son, Lett. Math. Phys. **4**, 93 (1980).
[8] S. Kochen and E. P. Specker, J. Math. Mech. **17**, 59 (1967).
[9] R. W. Spekkens, Phys. Rev. A **71**, 052108 (2005).
[10] M. Banik, S. S. Bhattacharya, S. K. Choudhary, A. Mukherjee, and A. Roy, Found. Phys. **44**, 1230 (2014).
[11] E. Schrödinger, Math. Proc. Cambridge Philos. Soc. **32**, 446 (1936).
[12] H. M. Wiseman, S. J. Jones, and A. C. Doherty, Phys. Rev. Lett. **98**, 140402 (2007).
[13] N. Harrigan and R. W. Spekkens, Found. Phys. **40**, 125 (2010).
[14] M. S. Leifer and O. J. E. Maroney, Phys. Rev. Lett. **110**, 120401 (2013).
[15] A. Peres, Phys. Lett. A **151**, 107 (1990); N. D. Mermin, Phys. Rev. Lett. **65**, 3373 (1990).
[16] A. Grudka *et al.* Phys. Rev. Lett. **112**, 120401 (2014); S. Yu and C. H. Oh, *ibid.* **108**, 030402 (2012); A. Cabello, arXiv:1201.0374.
[17] R. W. Spekkens, D. H. Buzacott, A. J. Keehn, B. Toner, and G. J. Pryde, Phys. Rev. Lett. **102**, 010401 (2009).
[18] R. Kunjwal and R. W. Spekkens, Phys. Rev. Lett. **115**, 110403 (2015).
[19] M. D. Mazurek *et al.*, arXiv:1505.06244.
[20] M. F. Pusey, arXiv:1506.04178.
[21] J. Barrett, Phys. Rev. A **75**, 032304 (2007).
[22] P. Janotta, C. Gogolin, J. Barrett, and N. Brunner, New J. Phys. **13**, 063024 (2011).
[23] S. Wiesner, SIGACT News **15**, 78 (1983).
[24] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani, arXiv:quant-ph/9804043.
[25] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani, J. Assoc. Comput. Mach. **49**, 496 (2002).
[26] P. Busch, M. Grabowski, and P. J. Lahti, *Operational Quantum Physics* (Springer-Verlag, Berlin, 1997).
[27] H. Barnum and A. Wilce, arXiv:0908.2352.
[28] C. Pfister, arXiv:1203.5622.
[29] N. Stevens and P. Busch, Phys. Rev. A **89**, 022123 (2014).
[30] I. Namioka and R. R. Phelps, Pac. J. Math. **31**, 469 (1969).
[31] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevA.92.030103 for the proof of optimality of the 2-bit POM protocol.
[32] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Rev. Mod. Phys. **81**, 865 (2009).
[33] M. Pawlowski *et al.*, Nature (London) **461**, 1101 (2009); M. Navascués and H. Wunderlich, Proc. R. Soc. London, Ser. A **466**, 881 (2009); J. Oppenheim and S. Wehner, Science **330**, 1072 (2010); M. Banik, Md. R. Gazi, S. Ghosh, and G. Kar, Phys. Rev. A **87**, 052125 (2013); A. Cabello, *ibid.* **90**, 062125 (2014).
[34] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevA.92.030103 for a complete proof of Proposition 3.
[35] C. H. Bennett, D. P. DiVincenzo, P. W. Shor, J. A. Smolin, B. M. Terhal, and W. K. Wootters, Phys. Rev. Lett. **87**, 077902 (2001).
[36] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevA.92.030103 for an entanglement-assisted scheme in QM.
[37] R. W. Spekkens, Phys. Rev. A **75**, 032110 (2007).
[38] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevA.92.030103 for an example of a steerable toy bit theory not useful for POM games.
[39] A. Chailloux, I. Kerenidis, S. Kundu, and J. Sikora, arXiv:1404.5153.