Operational foundations for complementarity and uncertainty relations

Debashis Saha* and Michał Oszmaniec†
Institute of Theoretical Physics and Astrophysics, National Quantum Information Centre, Faculty of Mathematics, Physics, and Informatics, University of Gdansk, Wita Stwosza 57, 80-308 Gdańsk, Poland and Center for Theoretical Physics, Polish Academy of Sciences, Al. Lotników 32/46, 02-668 Warsaw, Poland

Lukasz Czekaj‡
Faculty of Applied Physics and Mathematics, National Quantum Information Centre, Gdańsk University of Technology, 80-233 Gdańsk, Poland

Michał Horodecki§ and Ryszard Horodecki‖
Institute of Theoretical Physics and Astrophysics, National Quantum Information Centre, Faculty of Mathematics, Physics, and Informatics, University of Gdansk, Wita Stwosza 57, 80-308 Gdańsk, Poland

(Received 29 November 2019; accepted 23 March 2020; published 6 May 2020)

The so-called preparation uncertainty that occurs in the quantum world can be understood well in purely operational terms, and its existence in any given theory, perhaps differently than in quantum mechanics, can be verified by examining only measurement statistics. Namely, one says that uncertainty occurs in some theory when for some pair of observables, there is no preparation that would exhibit deterministic statistics for both of them. However, the right-hand side of the uncertainty relation is not operational anymore if we do not insist that it is just the minimum of the left-hand side for a given theory. For example, in quantum mechanics, it is some function of two observables that must be computed within the quantum formalism. Also, while joint nonmeasurability of observables is an operational notion, the complementarity in Bohr's sense (i.e., in terms of information needed to describe the system) has not yet been expressed in purely operational terms. Here we propose a general operational framework that provides answers to the above issues. We introduce an operational definition of complementarity and further postulate that complementary observables have to exhibit uncertainty, which means that we propose to put the (operational) complementarity as the right-hand side of the uncertainty relation. In particular, we identify two different notions of uncertainty and complementarity for which the above principle holds in the quantum-mechanical realm. We also introduce postulates for the general measures of uncertainty and complementarity. In order to define quantifiers of complementarity we first turn to the simpler notion of independence that is defined solely in terms of the statistics of two observables. Importantly, for clean and extremal observables—i.e., ones that cannot be simulated irreducibly by other observables—any measure of independence reduces to the proper complementary measure. Finally, as an application of our general framework we define a number of complementarity indicators and show that they can be used to state uncertainty relations. One of them, assuming some natural symmetries, leads to the Tsirelson bound for the Clauser-Horne-Shimony-Holt (CHSH) inequality. Lastly, we show that for a single system a variant of information causality called the information content principle, under the above symmetries, can be interpreted as an uncertainty relation in the above sense.

DOI: 10.1103/PhysRevA.101.052104

I. INTRODUCTION

Uncertainty and complementarity are landmark features of quantum mechanics and have been investigated since its inception almost a century ago. The concept complementarity captures the fact that in quantum mechanics two quantum observables cannot be measured simultaneously and hence supply "independent" pieces of information about a physical systems [1]. The uncertainty principle, proposed for the first time by Heisenberg, on the other hand, limits the precision of outcome statistics of two complementary observables, such as position and momentum [2]. Uncertainty relations are quantitative emanations of the uncertainty principle and play predominant role in the conceptual [3] and mathematical foundations of quantum theory [4–8]. Importantly, with the advent of quantum information, uncertainty relations found also practical applications in fields such as entanglement detection [9,10], quantum steering [11], as well as randomness generation and quantum cryptography [12].

*saha@cft.edu.pl
†michal.oszmaniec@gmail.com
‡jasiek.gda@gmail.com
§fizmh@ug.edu.pl
‖fizrh@ug.edu.pl

Despite the great success of the research effort concerning uncertainty relations, this line of research is inherently restricted to quantum formalism and so the notions of complementarity, uncertainty, and uncertainty relations have not been much explored outside quantum theory. Uncertainty itself is defined pretty operationally, and it was explored in a more general setup than the quantum one (see, e.g., [13–16]). The uncertainty relations were also considered in those papers. However, the right-hand sides of these relations were not expressed in operational terms. Also, while the issue of joint nonmeasurability (incompatibility) was explored outside of quantum-mechanical formalism [17–21], the complementarity of observables, understood in Bohr's sense, seems not to have been investigated so far in operational terms (apart from the approach in which complementarity is simply understood as just the minimum of the left-hand side of the uncertainty relation; cf. [3]).

This article aims to change this state of affairs. By defining complementarity in purely operational fashion, i.e., solely in terms of the measurement statistics of a given theory, we are able to obtain an operational form of the uncertainty relation—in which both sides of inequalities are some functions of just statistics of observables—without referring to the internal formalism of the theory.

The complementarity should be associated with the independent information that can be obtained from two different observables which cannot be measured jointly. This notion of complementarity is motivated by Bohr's own views concerning this concept. In one of the letters to Einstein [22] Bohr describes complementarity as follows [23]: "Evidence obtained under different experimental conditions cannot be comprehended within a single picture, but must be regarded as complementary in the sense that only the totality of the phenomena exhaust the possible information about the objects."

Furthermore, we postulate, inspired by quantum mechanics, that in reasonable physical theories uncertainty should be present for all complementary observables (i.e., we identify the right-hand side of the uncertainty relation with complementarity). In other words, uncertainty should be regarded as a price that we pay for complementarity of two observables. This fundamental trade-off we refer to as the *uncertainty principle*. On the other hand, for maximally informative measurements uncertainty should also imply complementarity. Such a relation can be called the *reverse uncertainty relation*.

Let us now outline the structure of this work. First, in Sec. II, we present the general operational framework in which we cast the concepts of uncertainty and complementarity. Then, in Sec. III we present the connections between various kinds of complementarity and uncertainty in quantum theory. Importantly, we observe that in quantum mechanics there are three different notions of uncertainty. We observe that for two of them there exist different but operationally well-motivated notions of complementarity that can be used to formulate *uncertainty principles*. We propose that in all reasonable physical theories the analogs of the aforementioned uncertainty principles should hold. In Sec. IV we argue that often complementarity of two clean and extremal observables (i.e., ones that cannot be simulated irreducibly by other observables) can be defined solely in terms of their output statistics. This is a great simplification as it allows us to (in some cases) discuss complementarity without any direct reference to the formalism or the structure of a particular theory. In Sec. V we present the intuitive exposition of our ideas in the case of dichotomic observables. This simple setting allows for a nice geometrical interpretation of our ideas concerning uncertainty, complementarity, and the uncertainty principle. After the first part of the paper, which has a rather introductory and conceptual flavor, in Sec. VI we give an overview and motivation for the technical results given in the latter part of the paper. Sections VII and VIII present our postulates for measures of uncertainty as well as complementarity and independence, respectively. In Sec. IX we propose a number of concrete measures of uncertainty and complementarity that are motivated either by the operational or geometrical considerations. Finally, in Sec. X we use some of these measures to state (apparently new) quantitative uncertainty relations valid in quantum mechanics. We also apply one such relation, together with the no-signaling assumption, to obtain Tsirelson's bound in the Clauser-Horne-Shimony-Holt (CHSH) inequality. We conclude the paper in Sec. XI, where we state a number of open problems and directions of further research. We also include appendices containing proofs of certain technical statements given in the main text.

II. FRAMEWORK AND NOTATION

First, we give a survey of notations and concepts used by us in this work. We will work in the framework of operational theories [24,25]. An operational theory consists of preparations P (belonging to the set $\mathcal{P}$) and measurements M (belonging to the set $\mathcal{M}$). An operational theory describes the statistics in a "prepare and measure" scenario, in which a system is prepared using a preparation procedure P and measured using a measurement device (observable) M. Then, the outcome k occurs with the probability $q_M(k|P)$. We will use the notation $\mathbf{q}_M(P) \equiv (q_M(1|P), \dots, q_M(n|P))$ to denote the vector of outcome statistics, when a preparation P is measured by a measurement M (n is the number of outcomes of M). From now on, for the sake of simplicity we will focus on the case of two measurements (observables) X, Y. *A priori* in the operational theory X and Y cannot be measured jointly; i.e., one does not have access to the joint probability distribution of observing values of *both* X and Y in a single experiment. Hence, in what follows we will be interested in distributions possible to obtain when measuring *either* of the observables X or Y. Therefore, for a given preparation procedure P, the object of interest is then the vector of probability distributions

$$\mathbf{q}(P) := (\mathbf{q}_X(P), \mathbf{q}_Y(P)), \quad (1)$$

where we dropped the dependence of $\mathbf{q}(P)$ on observables X, Y in order to keep the notation compact. As the preparation P varies we obtain different probability distributions $\mathbf{q}_{X,Y}(P)$ and consequently different vectors $\mathbf{q}(P)$. We denote the convex set of all allowed vectors $\mathbf{q}(P)$ by $S_{X,Y}$. The set $S_{X,Y}$ shall be called the *statistics set for X and Y*, or in short the *statistics set*. Thus, $S_{X,Y}$ is embedded in the Cartesian product of two simplices $S := \Delta_n \times \Delta_n$ (see Fig. 1).

Note that the set $\mathcal{P}$ can be always assumed to be convex as one can always formally define the mixture of two different preparations via the mixture of the corresponding probability

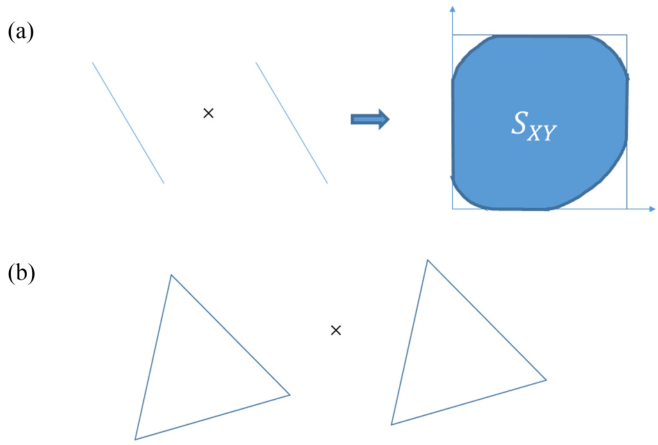

FIG. 1. The statistics set is embedded into the Cartesian product of two simplices $S = \Delta_n \times \Delta_n$. (a) Dichotomic observables ($n = 2$): Simplices are one-dimensional and the axes represent probabilities of a single outcome for each observable. Their Cartesian product is a square, and $S_{X,Y}$ is its convex subset. (b) Observables with three outputs ($n = 3$) give rise to two-dimensional simplices. Their Cartesian product and the set $S_{X,Y}$ cannot be visualized.

distributions for all measurements $X \in \mathcal{M}$. Operationally this corresponds to choosing between two preparation procedures P_1 and P_2 by the result of tossing of a biased coin with probability, say, $(\alpha, 1-\alpha)$. The output statistics of the resulting preparation P is convex linear, i.e.,

$$\forall X \in \mathcal{M},\ \mathbf{q}_X(P) = \alpha \mathbf{q}_X(P_1) + (1-\alpha)\mathbf{q}_X(P_2), \tag{2}$$

and therefore the statistics set $S_{X,Y}$ is convex. Similarly, one can perform a convex mixture of two different measurements $X_1, X_2 \in \mathcal{M}$ such that the output statistics for all preparation of the resulting measurement X is a convex combination of the corresponding probability distribution, i.e.,

$$\forall P \in \mathcal{P},\ \mathbf{q}_X(P) = \alpha \mathbf{q}_{X_1}(P) + (1-\alpha)\mathbf{q}_{X_2}(P). \tag{3}$$

Remark. Connecting to the standard quantum formalism: in quantum theory preparations P are simply quantum states whereas measurements (observables) M are simply allowed quantum-mechanical measurements.

The main aim of this work is to define and study the joint uncertainty [26], complementarity, uncertainty relations, and uncertainty principle in terms of the observed statistics $\mathbf{q}(P)$ and the allowed statistics set $S_{X,Y}$.

In what follows we will need a couple more concepts related to classical manipulation and simulation of observables in general theories. See [27,28] for the basic definitions in quantum mechanics, [29–31] for application in quantum information, and a recent work [32] for the extension to the realm of general probabilistic theories.

Definition 1 (simulation of observables). We say that observable X can simulate observable Y (denoted as $X \to Y$) when there exists a stochastic channel Λ such that if we apply the channel to outputs of the observable X, then for any preparation P, the obtained statistics is the same as the statistics of the outputs of Y for that preparation. This stochastic channel is equivalent to classical postprocessing of the outputs of the observables.

Formally, for $X \to Y$ there exists a stochastic map Λ such that $\mathbf{q}_Y(P) = \Lambda \mathbf{q}_X(P)$, *simultaneously*, for all preparations P. For instance, if Y is a coarse-grained version of X, then $X \to Y$.

Definition 2 (clean observables). An observable X is called *clean* if for any Y such that $Y \to X$, also $X \to Y$.

In other words, a clean observable is an observable that cannot be simulated in an irreducible manner to another observable in the theory.

Definition 3 (sharp observables). An observable X is called *sharp* if for any output there exists a P which gives this output with probability 1.

Definition 4 (extremal observables). We say that an observable X is extremal if the statistics of the outputs cannot be obtained by a convex mixture of two distinct measurements simultaneously for all preparations.

In quantum theory, any projective measurement composed of projectors of arbitrary rank is a sharp as well as extremal observable. Moreover, certain positive-operator-valued measures (POVMs) such as a "symmetric, informationally complete" POVM are examples of extremal observables, whose statistics cannot be expressed as a convex mixture of other measurements, while any projective measurement with only rank-1 projectors is a clean observable in quantum theory.

III. UNCERTAINTY, COMPLEMENTARITY, AND UNCERTAINTY RELATIONS

A. Preparation uncertainty relation and complementarity

Let us start with the formal definition of (preparation) uncertainty of two observables X, Y.

Definition 5 (joint preparation uncertainty). We say that a preparation P exhibits joint preparation uncertainty for observables X, Y if at least one of the distributions $\mathbf{q}_X(P)$, $\mathbf{q}_Y(P)$ is not deterministic.

In quantum mechanics the preparation uncertainty relation (PUR) [3] refers to the situation in which for two quantum-mechanical observables X, Y there exists no preparation (state) P for which both X and Y have well-defined values. Typically, the PUR has the form

$$\mathrm{U}_{X,Y}(P) \geqslant \mathrm{C}_{X,Y}, \tag{4}$$

where $\mathrm{U}_{X,Y}(P)$ is some measure of joint uncertainty of X and Y on a preparation P and $\mathrm{C}_{X,Y}$ is the quantity depending on X and Y. Often, the right-hand side of (4) is identified with the measure of complementarity of observables X and Y. Our goal is to propose a framework allowing us to consider the preparation uncertainty relation in any theory. Therefore, both sides of the PUR should have an operational interpretation, i.e., should depend only the observed statistics rather than on the formalism of the particular theory.

Currently, in quantum mechanics the right-hand side of the PUR is typically not defined operationally. Namely, it usually refers explicitly to the mathematical structure of quantum mechanics rather than to the observed statistics. For example, in the Kennard-Robertson uncertainty relation [33,34] $\mathrm{C}_{X,Y}$ depends on the commutator $[X, Y]$. Also, in Deutsch [35] and Maassen-Uffink [36] the entropic UR $\mathrm{C}_{X,Y}$ is a function of the maximal overlap of eigenvectors of the involved observables. Let us note that in quantum mechanics the right-hand side of (4) is nontrivial only for noncommuting observables. Such observables have a crucial feature that they *access pieces of information that cannot be obtained simultaneously*. In fact, this characteristic has been associated with complementarity already since the invention of quantum theory [1,22,37,38]. In this work we propose to define the notion of complementarity of two observables via impossibility of joint access to pieces of information obtained in the course of their measurements. This allows us to talk about complementarity in any physical theory. Importantly, our definition differs from the approach of [3], where complementarity is defined by the minimal value of uncertainty [the right-hand side of (4)] over all states allowed in the theory. This perspective, albeit operational, treats complementarity only as the quantifier of uncertainty of a theory. Our approach is that complementarity can be regarded as something positive: there is more information in the system than one observable, even the most fine-grained one, can access. This however, at least in quantum mechanics, comes with the price which takes the form of *uncertainty relations* (of various types that we discuss below). The existence of such price for the phenomenon of excess of information we shall postulate as a physical principle.

B. Complementarity and joint nonmeasurability

Let us start with the qualitative definition of complementarity.

Definition 6 (complementarity). We shall call two observables X, Y complementary if they are not jointly measurable; i.e., their statistics $\mathbf{q}_X(P)$, $\mathbf{q}_Y(P)$ cannot be obtained by classical postprocessing independently of the preparation P.

This definition is motivated by the following observation: if two observables are jointly measurable, this means that both pieces of information can be accessed by measuring a single observable. This would mean that the observables were simply not fine grained enough. Interestingly, this reasoning, in quantitative form, is itself an uncertainty relation, called the *measurement uncertainty relation* (MUR); quoting [39]: "Measurement uncertainty relations are quantitative bounds on the errors in an approximate joint measurement of two observables."

In quantum mechanics the two uncertainty relations—the MUR and PUR—are intimately related. Namely, the PUR can be nontrivial only for those observables for which the MUR holds. Here, we say that the PUR is nontrivial if it nontrivially restricts the statistics of the two observables; i.e., the right-hand side of (4) is nonzero.

Let us emphasize here that it is not always opposite: namely, even if observables are not jointly measurable (i.e., when we have a nontrivial MUR), the PUR may be still trivial. In other words, complementarity does not always enforce uncertainty. For example, when we have two observables that have a common eigenstate, but otherwise do not commute, we have no joint measurability, and the observables are still (though not fully) complementary but the PUR is trivial: the right-hand side of the PUR is zero, and there is no uncertainty. A basic example is given by these observables:

$$\begin{bmatrix} \sigma_x & 0 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} \sigma_z & 0 \\ 0 & 1 \end{bmatrix}. \tag{5}$$

Interestingly, even more drastic phenomena can happen. Consider two dichotomic projective measurements M and N in $\mathbb{C}^6$ (equipped with the standard basis $\{|i\rangle\}_{i=1}^6$) having the following effects,

$$\begin{aligned} M_1 &= |1\rangle\langle 1| + |3\rangle\langle 3| + |4\rangle\langle 4| , \\ M_2 &= |2\rangle\langle 2| + |5\rangle\langle 5| + |6\rangle\langle 6| , \\ N_1 &= |+\rangle\langle +| + |3\rangle\langle 3| + |5\rangle\langle 5| , \\ N_2 &= |-\rangle\langle -| + |4\rangle\langle 4| + |6\rangle\langle 6| , \end{aligned} \tag{6}$$

where $|\pm\rangle = (1/\sqrt{2})(|1\rangle \pm |2\rangle)$. It can be seen that even though the above measurements are not jointly measurable (because the states $|\pm\rangle\langle\pm|$ do not commute with the states $|1\rangle\langle 1|, |2\rangle\langle 2|$), there is *no uncertainty*. In fact for each pair of outcomes of M, N, there exists a state such that both the outcomes are certain. Thus, the statistics set $S_{N,M}$ is as big as possible and equals S, the Cartesian product of two one-dimensional simplices (see Fig. 1). Notice however that the above projective measurements are not clean (see Definition 2) since they can be obtained as coarse grainings of fine-grained (rank-1) projective measurements in $\mathbb{C}^6$. In what follows we will show that in quantum mechanics (suitably understood) joint nonmeasurability indeed implies (suitably understood) uncertainty, but only for clean and extremal observables.

C. Three types of uncertainty and complementarity

The above discussion shows that joint nonmeasurability may seem to be not a good candidate for the right-hand side of (4). Fortunately, there is an extension of the PUR: the exclusion principle proposed by Hall [40]. While the original Hall principle is still trivial for observables that share a common eigenstate, its natural extension conjectured in [41] and proved in [42] is nontrivial, whenever observables do not commute. The exclusion principles are quantified in a particular manner (via mutual information). We would like to avoid using any particular quantifiers as at the moment we are only interested in the question of whether there is uncertainty, or not, and whether there is information exclusion or not. In what follows we present the qualitative definitions of uncertainty and exclusivity that avoid usage of any quantifiers.

Definition 7 (traditional uncertainty). Two observables X, Y exhibit nonzero (preparation) uncertainty if for arbitrary preparation P, their statistics $\mathbf{q}_X(P)$, $\mathbf{q}_Y(P)$ are never both deterministic at the same time.

Definition 8 (information exclusion). Consider two observables X, Y with d outcomes. We say that they have *information exclusion* if there does not exist a d-element set of preparations P_i such that each of the preparations gives fully predictable outputs for both observables, and different preparations lead to a different outputs (statistics).

From now on we can operate solely on a qualitative level. In quantum mechanics, whenever sharp and clean measurements (i.e., projective measurements with one-dimensional projections) are not jointly measurable (equivalently, they do not commute [43]), they lead to the nontrivial information exclusion principle; *ergo* complementarity of two observables always imposes the nontrivial exclusion principle on those observables. For a formal proof see Lemma 1 in Appendix A.

TABLE I. Uncertainty and complementarity: Incomplete.

Uncertainty	Complementarity
Information exclusion.	Associated with joint nonmeasurability.
Traditional.	?

Recall that for nonclean observables, this is not true, as shown by measurements given in Eq. (6). Note that in quantum theory sharp and clean observables are extremal too. Thus, we have the following: *In quantum mechanics for clean observables complementarity implies information exclusion.*

As said, we cannot replace in this sentence "information exclusion" with "uncertainty." Thus we obtain the picture illustrated by Table I, where we have one space to fill: some version of complementarity, which would imply traditional uncertainty.

Now we would like to fill the space. Let us note that if we coarse-grain the observables from the example given in Eq. (5), by choosing not to distinguish between the two outcomes of σ_x and the same σ_z, then the new observables will become trivial, having no complementarity and no uncertainty. This prompts us to consider a stronger version of complementarity, which can be called *full complementarity*.

Definition 9 (full complementarity). We say that two observables X, Y are fully complementary when after arbitrary coarse graining (apart from the trivial one, where none of the outcomes are distinguished) the observables still remain jointly nonmeasurable.

Clearly, such stronger complementarity *implies uncertainty in the traditional form for projective measurements* (this follows from Lemma 1 in Appendix A). However, let us consider the following example:

$$\begin{bmatrix} \sigma_x & 0 \\ 0 & \sigma_x \end{bmatrix}, \quad \begin{bmatrix} \sigma_z & 0 \\ 0 & \sigma_z \end{bmatrix}. \tag{7}$$

The above two observables do not exhibit full complementarity, yet they are uncertain. Thus, this notion is a bit too strong to be put in the table on the same level as traditional uncertainty. At first glance, such a strong notion of complementarity should be associated with the following strong version of uncertainty, which, to our knowledge has not been examined so far.

Definition 10 (strong preparation uncertainty). We say that two observables X, Y exhibit strong (preparation) uncertainty when they remain uncertain after any nontrivial coarse graining. In other words, it is impossible to find a preparation P such that $\sum_{i\in I} q_X(i|P) = \sum_{j\in J} p_Y(j|P) = 1$, for some nontrivial subsets I, J of the output spaces of X and Y, respectively.

Remark. It is also possible to define a strong information exclusion. Namely, we say that observables X, Y exhibit strong exclusion when after any coarse graining they still exhibit information exclusion. Interestingly, in quantum mechanics, the two notions become equivalent; however, in general (for some weird theory), they may be distinct.

Somewhat counterintuitively, it turns out that in quantum mechanics full complementarity does not imply full uncertainty, even for clean and extremal measurements (see Appendix B for a concrete counterexample in five dimensions). Therefore, in quantum mechanics full complementarity and strong uncertainty will not give rise to an uncertainty-like principle. In turns out that the version of complementarity that implies traditional uncertainty (for clean observables) is the following intermediate version of complementarity, which we shall call *single-outcome* complementarity. The proof is given in Lemma 1 in Appendix A.

TABLE II. Uncertainty and complementarity: Full picture.

Uncertainty	Complementarity
Information exclusion.	Associated with joint nonmeasurability.
Traditional.	Associated with single-outcome joint nonmeasurability.

Definition 11 (single-outcome complementarity). We say that two d-outcome observables X, Y exhibit single-outcome complementarity when, after coarse grainings that preserve one outcome and glue the remaining $d-1$ outcomes, the resulting dichotomic observables are still jointly nonmeasurable.

Summarizing, for quantum mechanics we have obtained the full picture, as shown in Table II. Let us remark that the notion of *coexistence* of two observables [44], defined as jointly measurable for all possible binarizations of their outcomes, implies *single-outcome* jointly measurable. And *single-outcome* jointly measurable implies weak coexistence, that is, jointly measurable under some binarizations [45].

D. Uncertainty principle as a physical postulate

Motivated by the analysis presented in the preceding part, we have found candidates for the right-hand sides of the general uncertainty relation (4). These will be one of the variants of complementarity, depending on what type of uncertainty we will put to the left-hand side. The implications between our notions, both those that hold by definition as well as those postulated as (qualitative) uncertainty relations, are depicted in Fig. 2. We also show in the figure the pair

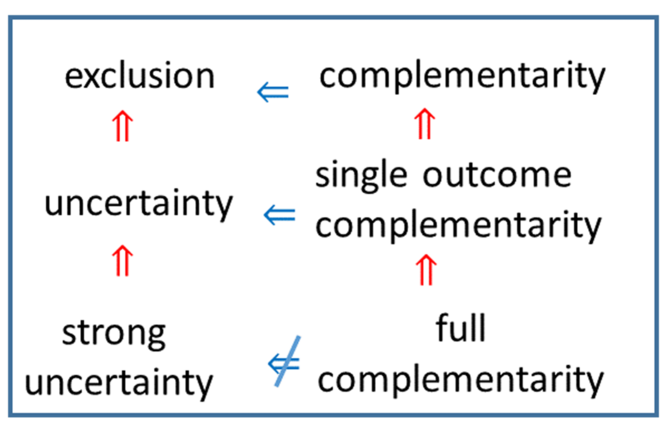

FIG. 2. The red implications hold by definition and, thus, hold in any theory. The blue ones hold for sharp projective observables in quantum theory. We further postulate the blue implications for clean, sharp, and extremal observables in physical theories.

strong uncertainty versus full complementarity, pointing out that the implication does not hold. Recall that in quantum mechanics, relations between the two kinds of uncertainty and complementarity given in Fig. 2 hold only for fine-grained projective measurements. These measurements are clean and extremal observables (see Definition 2) and we postulate the relation between uncertainty and complementarity only for clean-extremal measurements.

Postulate (uncertainty principle). In physical theories observables which are complementary, clean, and extremal necessarily exhibit uncertainty.

In other words, in any theory lack of joint measurability for clean and extremal observables must imply uncertainty. That is, the existence of the uncertainty principle can be also understood as a price for the excess of information provided by complementary observables: *The uncertainty principle states that complementarity has a price—which is uncertainty.*

Remark. Let us emphasize that while uncertainty is present only in the quantum world, and not in the classical one, the uncertainty principle holds both in quantum and classical theory: In the classical case it holds because there is no complementarity, and therefore the "price" is zero.

Remark. In this work we will be mostly interested in sharp and extremal observables, as nonsharp or nonextremal observables are themselves uncertain, and the uncertainty is not related to complementarity, but just comes from some form of *a priori* epistemic restrictions. Note however that the existence of nonsharp or nonextremal measurements *does not* contradict the uncertainty principle.

Later in this paper we shall pave the way to quantify uncertainty and complementarity, aiming to grasp the above principle quantitatively. At this moment let us informally state the general form of uncertainty relations.

Definition 12 [preparation uncertainty relation (PUR)]. The general preparation uncertainty relation is an inequality of the following form,

$$\mathrm{U}_{X,Y}(P) \geqslant f^{\uparrow}(\mathrm{C}_{X,Y}), \tag{8}$$

where $\mathrm{U}_{X,Y}(P)$ is a measure of (joint) uncertainty of X and Y, $\mathrm{C}_{X,Y}$ is some indicator of complementarity of observables X, Y (see Sec. VII for the properties that these quantities should satisfy), and $f^{\uparrow}$ is a nondecreasing function whose range is non-negative. The form of these two functions depends on the particular measures of complementarity and uncertainty used.

So far we have mostly talked about the negative aspect of complementarity (joint nonmeasurability); however, as we have mentioned, it is strictly connected with a positive aspect of complementarity: because of joint nonmeasurability, the observables reveal more information than possible by means of a single observable. Further in the paper we will provide examples of quantifiers of uncertainty that would reflect this point of view.

Remark. Let us emphasize that the notion of complementarity we propose differs from the one considered in [14]: "... two measurements are complementary if the second measurement can extract no more information about the preparation procedure than the first measurement and visa versa. We refer to this as information complementarity. Note that quantum mechanically, this does not necessarily have to do with whether two measurements commute. For example, if the first measurement is a complete von Neumann measurement, then all subsequent measurements gain no new information than the first one whether they commute or otherwise." We see that the authors consider *sequential* measurements, and that their definition incorporates the process of disturbing the state by measurement. In our paper we restrict ourselves to the typical scenario of preparation uncertainty relations, where there are no sequential measurements, and our complementarity is built-in in such a paradigm.

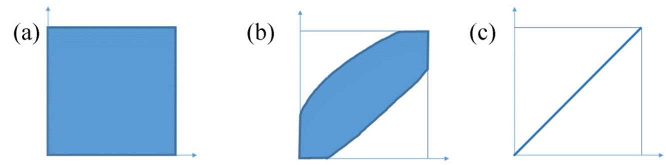

FIG. 3. The statistics set for (a) most independent observables, (b) intermediate case, and (c) the same observables.

E. Reverse uncertainty relations

One can also ask, How about the inverse relation, where complementarity would imply uncertainty? We may consider the following definition:

Definition 13. The reverse PUR is the following implication: nonzero uncertainty implies nonzero complementarity. That is, uncertainty cannot occur if observables are not complementary to some extent.

Note that while the uncertainty principle does not hold in arbitrary theory, and we want to propose it to be a postulate for legitimate theories, the above *reverse PUR* is expected to hold for all sharp and clean pairs of observables. In Sec. X C we present the result which says that the reverse PUR holds for binary, sharp, and clean outcomes for any theory. We give there the quantitative form of such reverse PUR. In quantum mechanics, it is easy to see that the reverse PUR holds qualitatively for the pair exclusion complementarity; i.e., exclusion implies complementarity.

IV. COMPLEMENTARITY FROM STATISTICS SET AND INDEPENDENCE

In the previous section, while discussing how to make uncertainty relations operational, we have put emphasis on the connection between complementarity and the impossibility of joint measurement. Yet, one should also embrace the positive aspect of complementarity: it is the surplus of information provided by two (or perhaps more) observables. In this section we would like to describe how one can quantify such excess in arbitrary theory.

Consider a very simple theory: it has just two dichotomic observables X and Y, and all possible pairs of distributions are allowed (i.e., for any pair of distributions there exists a preparation that gives rise to these distributions, via measurement of our observables). The statistics set $S_{X,Y}$ is therefore the full square (see Fig. 3). Clearly, each of them brings completely independent information, and these two pieces of information cannot be acquired in any other way. Thus the two observables are maximally complementary.

Suppose that the set $S_{X,Y}$ shrinks a bit toward one of the diagonals. The observables become correlated, although there is no joint distribution. Namely, measuring any of them does not bring a lot of new information, compared to the information already provided by the measurement of the other one. This is clearly visible in the extreme example, when the set $S_{X,Y}$ is just the diagonal and the observables are identical. Thus, *the more the set shrinks, the less it is complementary*. Since our two observables are the only ones in the theory, the complementarity is solely a function of the statistics set $S_{X,Y}$, i.e., $\mathrm{C}_{X,Y} = C(S_{X,Y})$. Moreover, it should be intuitively monotonic under inclusions; i.e., if $S_{X,Y} \subset S_{X'Y'}$ then $C(S_{X,Y}) \leqslant C(S_{X',Y'})$. To summarize: if X and Y are the only observables in the theory, complementarity can be identified with their "independence," which can be intuitively deduced from the statistics set.

The problem becomes more complicated when there are other observables in the theory. To see this, consider a quite opposite situation—two classical bits. X measures one bit, and Y measures the other. The set $S_{X,Y}$ is the same—again square. But complementarity vanishes, as the information can be accessed by refined observable with four outcomes—the two-bit observable. Thus, for observables that are not clean the statistics set does not tell us anything about complementarity.

Similarly, the statistics set of nonextremal observables does not capture complementarity. Suppose two observables X_1 and X_2 are not complementary with observable Y separately. We naturally expect that complementarity between Y and another observable X which is realized by some convex mixture of observables X_1, X_2 is also zero. However, in general "independence" does not satisfy this feature. We provide an example in Appendix C.

Therefore, in what follows we limit ourselves to clean and extremal observables. We can now ask again whether independence $\mathrm{Ind}(S_{X,Y})$ (the intuitively defined function of the statistics set $S_{X,Y}$, as elaborated above) is related to complementarity (joint nonmeasurability). Or more concretely, can we infer complementarity looking solely at the statistics set for two clean and extremal observables? By definition, for clean and extremal observables there does not exist any set of observables that might reproduce two observables exactly. If one observable can simulate the other one (see Definition 1) the statistics set has zero measure. Hence, if the set $S_{X,Y}$ is a bit thicker than just the diagonal, this must imply that we have complementarity.

However, quantitatively we might still have the following situation: there exists a third observable that almost simulates our observables X and Y. And this observable would be able to acquire almost all the information; hence the independence of the observables would again mean just standard independence, and would not imply complementarity. In such a theory, the (approximate) joint measurability is not revealed in the statistics set. Note that in quantum mechanics this is not so. Consider, e.g., qubit observables. When they are complementary, the set is a circle. When they become more and more similar (*ergo* more and more jointly measurable) the statistics set shrinks toward a diagonal (see Fig. 4).

To summarize: for clean and extremal observables independence may not reflect complementarity, in a theory in which better and better joint measurability of clean observables does not imply that the observables converge to one another. Thus in general one should somehow connect two features: (i) how well observables can be simulated by a third one (which is a subject of the MUR); (ii) independence seen in the statistics set. And complementarity would be a function of those two features. This looks like a very ambitious program, and therefore for the purpose of this paper, we shall take a first step. Namely, we shall work out complementarity, which will work well in theories where approximate joint measurability (for clean and extremal observables) means that the observables are approximately the same. Thus, in the rest of the paper, we will assume that the statistics set of clean and sharp observables properly reflects the joint measurability features.

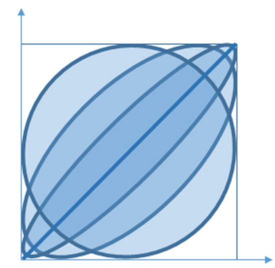

FIG. 4. The statistics sets for quantum binary observables. Circle is for most complementary (e.g., σ_x and σ_z, diagonal for both being σ_z).

Finally, we can define the complementarity through independence for arbitrary extremal observables as follows,

$$\mathrm{C}_{X,Y} := \min_{X':X'\to X} \min_{Y':Y'\to Y} \mathrm{Ind}(S_{X',Y'}), \tag{9}$$

where the minimum is taken over all observables X' and Y' that simulate X and Y, respectively. For nonextremal observables, we follow the convex-roof extension of the above definition, that is,

$$\mathrm{C}_{X,Y} = \min_{\{\alpha_i,X_i\}} \min_{\{\beta_j,Y_j\}} \sum_{i,j} \alpha_i \beta_j \mathrm{C}_{X_i,Y_j}, \tag{10}$$

where the minimum is taken over all possible decompositions of the observables X, Y to the extremal observables $\{X_i\}$, $\{Y_j\}$ with probability distribution $\{\alpha_i\}$, $\{\beta_j\}$. Importantly, this notion of complementarity reduces to independence for clean and extremal observables.

V. DICHOTOMIC OBSERVABLES: INTUITIVE PICTURE

In this part we focus exclusively on the case of dichotomic observables. This simplified setting allows for the appealing geometrical interpretations of the ideas presented in the preceding sections. As mentioned before, for two observables X, Y, each with two outputs, the simplices are just intervals, and the product of two simplices is a square. The set of $S_{X,Y}$ is some convex body within the square. Possible sets $S_{X,Y}$ are depicted in Fig. 5.

If both observables are sharp, i.e., for any outcome there exists a state that gives this outcome with probability 1, the set must touch each of the edges of the square. The examples of nonsharp observables are in Figs. 5(b) and 5(c). In Fig. 5(f) we have qubit observables of the form $X = \mathbf{n} \cdot \sigma$,

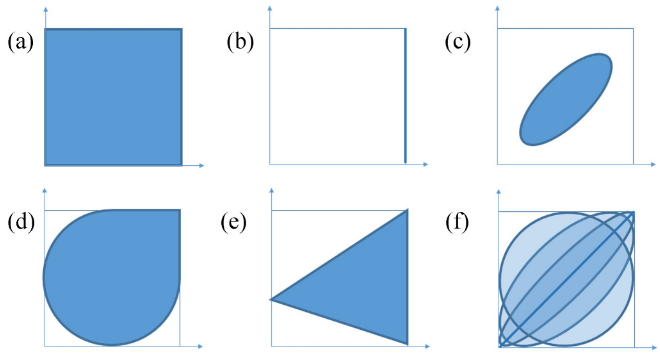

FIG. 5. Various sets $S_{X,Y}$. (a) $S_{X,Y}$ is equal to full square—the so-called "square bit." (b) One observable is completely noisy—reports no information. (c) Both observables are not sharp; i.e., there is no state that would give deterministic outcome for any of them. (d), (e) Both observables are sharp. (f) Quantum-mechanical observables.

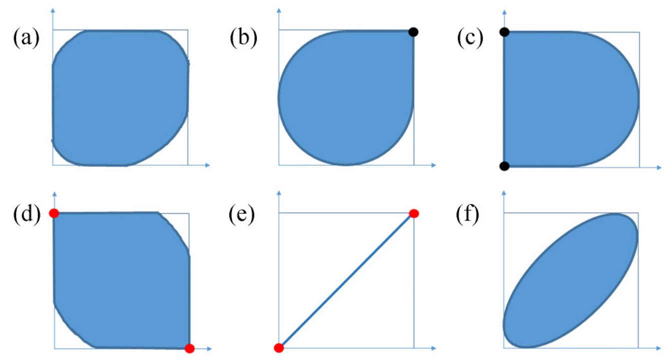

FIG. 6. Uncertainty for two-outcome observables. (a) No corner included, hence we have uncertainty for any state. (b) One corner included—represents preparation that has no uncertainty for both observables; exclusion still holds. (c) Two corners included, so for two preparations no uncertainty; still exclusion holds. (d) No uncertainty and no exclusion, since opposite corners are included. (e) Classical case (the same observables)—no uncertainty. (f) Generic quantum observables: both uncertainty and exclusion.

$Z = \sigma_z$, with $n_y = 0$ and $n_x^2 + n_z^2 = 1$. Depending on the angle between the vectors $\mathbf{n}$ and (0,0,1), we interpolate between (i) the classical case, where both observables are σ_z and the set $S_{X,Y}$ is just a line connecting opposite corners, and (ii) the most complementary case, where the set $S_{X,Y}$ constitutes a circle and observables are σ_x and σ_z. In the latter case the two observables are "mutually unbiased"; i.e., for any state that gives a deterministic outcome for one observable, it gives completely random output.

A. Independence and complementarity

Note that for two outcomes, there is no distinction between the three kinds of complementarity and independence presented in Sec. III C. This is because there are no nontrivial coarse-graining operations. Assuming that observables are clean and extremal, we can now identify complementarity and independence (see discussion in Sec. IV).

Square bit. For the states to be corners, both observables bring maximal and independent information. Clearly the square presents the richest statistics that can be obtained from two observables; therefore it has the largest possible independence among all sets $S_{X,Y}$.

Classical bit. The set $S_{X,Y}$ is just diagonal or antidiagonal. In the first case the second observable is just a copy of the first one, and in the second case, its negation. Here both observables report exactly the same information. *Ergo*, we have no independence.

Qubit. For observables $X = \mathbf{n} \cdot \sigma$, $Z = \sigma_z$, with $n_y = 0$ and $n_x^2 + n_z^2 = 1$, see Fig. 5(f) depending on angle between the vectors $\mathbf{n}$ and (0,0,1), we interpolate between the classical case, where observables are the same, and the most complementary case possible in quantum mechanics, where the set $S_{X,Y}$ constitutes a circle. This latter is the case in which two observables are "mutually unbiased"; i.e., for any state that gives a deterministic outcome for one observable, it gives completely random output. Note that this randomness is not a signature of complementarity. Exactly the same behavior occurs also for the square bit, where we can have states deterministic for both observables. Rather it should be regarded as *uncertainty*.

Generally, for dichotomic clean and extremal observables, whenever the statistics set is thick (i.e., not one-dimensional) we expect nonzero complementarity. In particular, the measures that we shall propose further, in the case of two outcomes, will all have this feature.

B. Uncertainty

The concept of uncertainty for dichotomic observables is illustrated in Fig. 6. The only preparations that give deterministic statistics for both observables correspond to corners of the square. The traditional uncertainty thus means that the set $S_{X,Y}$ does not include any corner. Exclusion means that the set does not include any pair of opposite corners. Thus, unlike in the case of complementarity, even for two outcomes, uncertainty does not reduce to one type: there can be a situation in which exclusion holds but there is no uncertainty; see Figs. 6(b) and 6(c). Clearly, strong uncertainty and traditional uncertainty collapse into one notion, since there is no nontrivial coarse graining for two outputs. Thus we are left with two types of uncertainty. Note that in quantum mechanics for two outcomes, at least qualitatively, there is no difference between the traditional uncertainty and exclusion.

Finally, note that in [13] theories were considered whose elementary systems exhibit the statistics set $S_{X,Y}$ described by the equation

$$\left(\frac{x-1}{2}\right)^p + \left(\frac{y-1}{2}\right)^p \leqslant 1 \tag{11}$$

for $p \geqslant 1$. For $p = 2$ it is a circle, i.e., the quantum case of maximally complementary observables. For $p \to \infty$ the set $S_{X,Y}$ tends to a full square.

C. Uncertainty principle

As said in Secs. III and IV, the preparation uncertainty principle says that there is a price for complementarity: namely, complementary observables have to be uncertain.

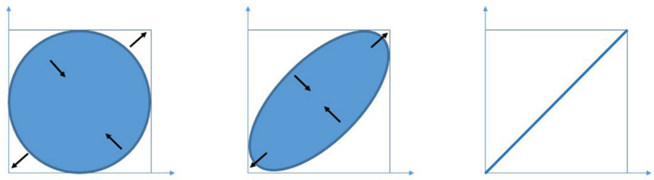

FIG. 7. Uncertainty principle in quantum case.

For two outcomes, the uncertainty principle says that whenever complementarity is nonzero, e.g., when the statistics set $S_{X,Y}$ is not one-dimensional, then the set does not contain corners. On a more quantitative level, the uncertainty principle says that the more complementarity we want, the larger must be the uncertainty. We see this in the quantum case: the more we want to be close to all four corners, the more we depart from the two original corners, which belonged to $S_{X,Y}$ in the case of the classical bit (i.e., when two observables were the same). We observe this in Fig. 7, where we show sets $S_{X,Y}$ for three values of the angle between observables.

Generally, uncertainty means that the set $S_{X,Y}$ is far from any of the corners. Complementarity means that S_{XY} is close to all the corners. Thus, the uncertainty principle says that when one wants to be close to any one of the two opposite corners, one cannot be close to the other opposite corner. Thus the uncertainty principle puts also bounds on complementarity itself: the maximal complementarity can be achieved only when uncertainty vanishes, but this is forbidden by the uncertainty principle.

VI. OUTLINE OF THE FURTHER RESULTS OF THE PAPER

In this section we will give the motivation and an overview of the results presented in the second half of the paper.

A. Quantifying independence and complementarity and proposing uncertainty relations

In the paper we shall propose some postulates that are measures of uncertainty (Sec. VII). They are just a modest updating of the postulates given in [7,26]. Then we propose postulates for measures of independence and complementarity in Sec. VIII.

Subsequently we shall propose some concrete measures of complementarity. Mostly we will concentrate on one of the types out of the three presented in Sec. III: the most basic one that does not involve coarse graining. We shall propose measuring by means of random access codes in Sec. IX A, by means of rescaling in Sec. IX B, and by means of preimages in Sec. IX C. *A priori* we might not be able to make from these the proper PUR, because in Table II they are in different rows. However, as already discussed in Sec. IV, for binary outcomes all complementarities coincide. We shall also propose uncertainty based on random access codes in Sec. IX A.

Now, having more or less compatible candidates for uncertainty and complementarity, one would like to build an uncertainty relation that might be imposed on all theories. Let us emphasize that we do not necessarily want the simple form of Eq. (4). We will be satisfied with any relation that will constrain uncertainty by complementarity.

One way of obtaining uncertainty relations to be imposed on physical theories is to find what relation between proposed uncertainty U and complementarity C is satisfied in quantum mechanics. An example of such a PUR will be the relation (54) between C and U built on the basis of rescaling.

B. Relation with information contents principle

Having proposed some understanding of what the uncertainty principle can mean in operational terms, it would be good to have a universal PUR that is not forcefully built to fit quantum mechanics. An example of a principle that holds in quantum mechanics even though it was not deliberately chosen to do so is information causality [46]. In [47] a version of information causality was proposed that differs mainly by putting emphasis on a single system, while information causality *a priori* deals with bipartite systems. It was called the information content principle (ICP). It represents a bound on random access codes for ensembles of states quantified by the mutual information. Therefore, qualitatively, it prevents maximal complementarity (if the latter is expressed by means of random access codes). In Sec. X D we will show that if the set S_{XY} is symmetric under rotation about $\pi/4$ as in the case of quantum mutually unbiased observables, then ICP turns out to be the Maassen-Uffink uncertainty relation for such observables. We also show that even with fewer symmetry assumptions, it still provides constraints for $S_{X,Y}$ which can play the role of the PUR; namely ICP prevents too much complementarity, if there is not much uncertainty.

C. Consequences of uncertainty relation for nonlocality

One of the interesting applications of the idea of operational uncertainty relations, which we will present in Sec. X E, is that they can put bounds on nonlocality. It is a ubiquitous problem of quantum-information theory to understand in operational terms what prevents quantum mechanics from being less nonlocal than would be possible if the only constraint were no signaling; see, e.g., [46,48,49] (in [50] the opposite direction was explored too: nonlocality and no signaling imply measurement uncertainty). Oppenheim and Wehner [14] attempt to understand why quantum mechanics is not maximally nonlocal; namely, they have made a crucial observation that the system that exhibits maximally nonlocal behavior, i.e., it violates the CHSH inequality up to its algebraic bound, exhibits no uncertainty.

Indeed, consider the CHSH inequality. Alice and Bob measure one of two observables A_1, A_2 and B_1, B_2. When Alice measures her observable A_1, and gets some outcome, she prepares the state on Bob's site. To maximize CHSH, Alice's outcome should be perfectly correlated with Bob's outcome, for any of his two observables. Thus the state of Bob's system, prepared by Alice's measurement and outcome, must give a deterministic answer to both his observables. Thus his observables cannot exhibit uncertainty. This suggests that it is uncertainty that bounds the nonlocality. However, there is a problem here: classical systems do not exhibit uncertainty, and yet still are not maximally nonlocal; even more, they are not nonlocal at all. Thus saying that uncertainty put bounds on nonlocality would be a very weak statement, as it would not provide any bound on nonlocality of classical systems, and in consequence could not capture the phenomenon of nonmaximal nonlocality of quantum mechanics.

The way out proposed in [14] was to involve also steering. To quote the authors: "... the degree of nonlocality of any theory is determined by two factors—the strength of the uncertainty principle, and the strength of a property called 'steering,' which determines which states can be prepared at one location given a measurement at another. ... For any physical theory we can thus consider the strength of nonlocal correlations to be a trade-off between two aspects: steerability and uncertainty." Some disadvantage of this approach is that it cannot be based only on statistics of observables in question. To verify the statement, the authors had first to find observables that are optimal for violation of the Bell inequality, and then for those observables optimize steering.

Here, we propose a different way out, one possible to spell out in operational terms. Namely we add to the word "uncertainty" another word "principle"; i.e., we say, "The uncertainty *principle* puts bounds on nonlocality." Since, as discussed above, the uncertainty principle holds for the whole quantum theory (unlike uncertainty, which appears only for specific observables), our statement implies also bounds on nonlocality for classical systems. We thus arrive at the following explanation of why quantum theory is not maximally nonlocal: *Quantum theory is not maximally nonlocal because of the uncertainty principle.*

Note that in [14] a stronger claim was made: namely, that uncertainty and steering not only bound the nonlocality, but they actually *determine* its value. This was later refuted in [15]. However, the weaker statement that uncertainty and steerability properties limit nonlocality is still meaningful. Also in our case, we are on the same level: we claim that the uncertainty principle puts bounds on nonlocality.

Here we will argue how the uncertainty principle bounds nonlocality for clean and extremal observables on a qualitative level. In Sec. X E we shall provide a quantitative picture, reproducing the Tsirelson bound. For binary outputs, notions of complementarity discussed in Sec. III all become the same. Thus, the uncertainty principle means qualitatively that complementarity implies uncertainty of any of three kinds. Now, for binary outcomes uncertainty means that the set $S_{X,Y}$ does not include any corner. Indeed if a corner belongs to $S_{X,Y}$, this means that there exists preparation, such that both distributions are deterministic. In Sec. X E we shall argue that from no signaling it follows that to have maximal violation of CHSH one needs two observables with set $S_{X,Y}$ being square. One can see this quickly in the following way: to violate CHSH maximally, one needs the so-called Popescu Rochrlich box. From its very definition it follows that after Alice's measurement, Alice prepares such states on Bob's side, and all four corners appear.

Now, we employ the uncertainty relation: since Bob's observables will have $S_{X,Y}$ being square, then complementarity is nonzero. However, the uncertainty principle says that then there must be uncertainty; i.e., the set cannot touch corners, and therefore cannot be a square. In short, the uncertainty principle rules out the square, and therefore CHSH cannot be maximally violated. A drawback of our approach is that it works only for clean and extremal observables. Observables that are not clean can have $S_{X,Y}$ to be square, without uncertainty—e.g., when one observable is one bit and the other is the other bit on the total system of two bits.

VII. POSTULATES FOR MEASURES OF UNCERTAINTY

In this part we give the postulates for measures of uncertainty for two observables. We shall not take the order from the weakest to the strongest (which would be as follows: exclusion, traditional uncertainty, strong uncertainty or exclusion) . Instead, we will begin with the most well known: uncertainty. Then, we will proceed with its immediate derivative—strong uncertainty—and end up with exclusion, which is the most complicated one.

A. Uncertainty

First, any measure of the joint uncertainty U of two observables (measurements) X and Y should depend on the observed statistics, in particular, the preparation procedure; i.e., we should have $\mathrm{U}(P) = \mathrm{U}(\mathbf{q}(P))$. Intuitively, the measure U should tell us to what extent it is impossible to have simultaneous knowledge about both X and Y for a given preparation P. We propose the following postulates for the measure of joint uncertainty (note that they are closely related to the postulates given in [7,26]).

(1) We assume $\mathrm{U}(\mathbf{q}(P)) \geqslant 0$ and $\mathrm{U}(\mathbf{q}(P)) = 0$ if and only if the distributions of X *and* Y giving rise to $\mathbf{q}(P)$ are deterministic. In other words $\mathbf{q}(P)$ is not located in the corner of the Cartesian product of two simplices; see Fig. 6.

(2) We assume that the $\mathrm{U}(\mathbf{q}(P))$ measure cannot decrease under doubly stochastic operations performed *independently* on outcomes of observables, i.e.,

$$\mathrm{U}((D_1, D_2)\mathbf{q}(P)) \geqslant \mathrm{U}(\mathbf{q}(P)), \tag{12}$$

for all doubly stochastic $n \times n$ matrices D_1 and D_2.

(3) The $\mathrm{U}(\mathbf{q}(P))$ measure cannot increase under coarse graining and permutations of outcomes. Formally,

$$\mathrm{U}((\Lambda_1^e, \Lambda_2^e)\mathbf{q}(P)) \leqslant \mathrm{U}(\mathbf{q}(P)) \tag{13}$$

for any *extremal* stochastic maps $\Lambda_{1,2}^e$.

(4) We assume that $\mathrm{U}(\mathbf{q}(P))$ cannot decrease under taking a mixture of preparations; i.e., U is concave with respect to the convex structure of preparations,

$$\mathrm{U}(\mathbf{q}(\alpha P_1 + (1-\alpha)P_2)) \geqslant \alpha \mathrm{U}(\mathbf{q}(P_1)) + (1-\alpha)\mathrm{U}(\mathbf{q}(P_2)), \tag{14}$$

for all $\alpha \in [0, 1]$.

(5) We assume that uncertainty cannot decrease for a mixture of measurements. Therefore, U is concave with respect to the convex structure of measurements, i.e.,

$$\mathrm{U}(\mathbf{q}_X(P), \mathbf{q}_Y(P)) \geqslant \alpha \mathrm{U}(\mathbf{q}_{X_1}(P), \mathbf{q}_Y(P)) + (1-\alpha)\mathrm{U}(\mathbf{q}_{X_2}(P), \mathbf{q}_Y(P)), \tag{15}$$

where the observable X is realized by the convex mixture of two observables X_1, X_2 with probability distribution $(\alpha, 1-\alpha)$.

Now *uncertainty of the statistics set*, $\mathrm{U}(S_{X,Y})$, is defined by the minimum U over all tuples of distributions in $S_{X,Y}$,

$$\mathrm{U}(S_{X,Y}) := \min_{x\in S_{X,Y}} \mathrm{U}(x). \tag{16}$$

(i) From concavity of $\mathrm{U}(\mathbf{q}(P))$, Eq. (14), it follows that the minimum in Eq. (16) is attained for the extremal points of $S_{X,Y}$.

(ii) The uncertainty measure possesses well-defined behavior under inclusion; i.e., for $S' \subset S$ we have

$$\mathrm{U}(S) \leqslant \mathrm{U}(S'). \tag{17}$$

(iii) It follows form postulate Eq. (12) that any uncertainty measure is invariant under all doubly stochastic operations whose inverse is also a doubly stochastic operation. For instance, uncertainty is invariant under all possible relabeling (or permutations) of the outcomes.

B. Strong uncertainty

We postulate any measure of strong (or full) uncertainty, which is denoted by $\mathrm{U}^f(\mathbf{q}(P))$, to be nonzero only if uncertainty is nonzero for all possible coarse graining of outcomes except the trivial one. Formally, $\mathrm{U}^f(\mathbf{q}(P)) = 0$ if there exists extremal maps Λ_1^e, Λ_2^e such that $\mathrm{U}((\Lambda_1^e, \Lambda_2^e)\mathbf{q}(P)) = 0$, where Λ_1^e, Λ_2^e corresponds to the all possible permutations and coarse graining except the trivial one.

Apart from that it also satisfies the postulates (12), (13), (14), and (15) of uncertainty.

C. Information exclusion

Here we list the postulates for any measure of information exclusion of $S_{X,Y}$.

(1) $\mathrm{E}(S_{X,Y}) \geqslant 0$ and $\mathrm{E}(S_{X,Y}) = 0$ if and only if for all outcome k, there exists a preparation, say P_k, such that

$$\mathbf{q}_X(k|P_k) = \tilde{\mathbf{q}}_Y(k|P_k) = 1, \tag{18}$$

where $\tilde{\mathbf{q}}_Y(k|P_k)$ is an arbitrary n-element permutation of $\mathbf{q}_Y(k|P_k)$, i.e., $\tilde{\mathbf{q}}_Y(k|P_k) = \mathbf{q}_Y(\pi(k)|P_k)$.

(2) $\mathrm{E}(S_{X,Y})$ cannot decrease under doubly stochastic operations performed *independently* of outcomes of observables, i.e.,

$$\mathrm{E}((D_1, D_2)S_{X,Y}) \geqslant \mathrm{E}(S_{X,Y}), \tag{19}$$

for all doubly stochastic $n\times n$ matrices D_1 and D_2. Here, $\mathrm{E}((D_1, D_2)S_{X,Y})$ denotes the allowed probability distribution in S obtained from the observed statistics $(D_1, D_2)\mathbf{q}(P)$.

(3) The $\mathrm{E}(S_{X,Y})$ measure cannot increase under coarse graining of outcomes. Formally,

$$\mathrm{E}\big((\Lambda_1^e, \Lambda_2^e)S_{X,Y}\big) \leqslant \mathrm{E}(S_{X,Y}) \tag{20}$$

for any *extremal* stochastic maps $\Lambda_{1,2}^e$.

(4) The $\mathrm{E}(S_{X,Y})$ measure possesses well-defined behavior under inclusion; i.e., for $S' \subset S$ we have

$$\mathrm{E}(S) \leqslant \mathrm{E}(S'). \tag{21}$$

(5) Exclusion cannot decrease under a convex mixture of measurements, i.e.,

$$\mathrm{E}(S_{X,Y}) \geqslant \alpha\mathrm{E}(S_{X_1,Y}) + (1-\alpha)\mathrm{E}(S_{X_2,Y}), \tag{22}$$

where the observable X is realized by the convex mixture of two observables X_1, X_2 with probability distribution $(\alpha, 1-\alpha)$.

VIII. POSTULATES FOR MEASURES OF INDEPENDENCE AND COMPLEMENTARITY

In this section we give the postulates for measures of independence and complementarity for two observables.

A. Independence

Recall that according to the notation introduced in Sec. II, $X \to Y$ means that for observables X, Y there exists a stochastic map Λ such that $\mathbf{q}_Y(P) = \Lambda\mathbf{q}_X(P)$, *simultaneously*, for all preparations P.

Now we propose that any measure of independence (Ind) should depend only on the statistics that can be possibly observed while measuring X or Y, that is, on the set $S_{X,Y}$. Here are our postulates for the measure of independence:

(1) We assume $\mathrm{Ind}(S_{X,Y}) \geqslant 0$ and that $\mathrm{Ind}(S_{X,Y}) = 0$ if $X \to Y$ or $Y \to X$.

(2) Any independence measure is invariant under independent relabeling of outcomes of X and Y, that is,

$$\mathrm{Ind}((\pi_1, \pi_2)S_{X,Y}) = \mathrm{Ind}(S_{X,Y}), \tag{23}$$

for all permutations $\pi_{1,2}$ of an n-element set. $(\pi_1, \pi_2)S_{X,Y}$ denotes the allowed region obtained from the observed statistics of $(\pi_1, \pi_2)\mathbf{q}(P)$.

(3) Independence is a "monotonic" function of S under inclusion; i.e., for $S' \subset S$ we have

$$\mathrm{Ind}(S) \geqslant \mathrm{Ind}(S'). \tag{24}$$

Remark. It might seem natural to require monotonicity under postprocessing, i.e., any stochastic map applied to outcomes of observables. However, it may happen that before processing observables are in the relation "→", i.e., one can simulate the other one, yet after some channel, they are no longer in the relation. Now, we require that independence is zero for observables that are in the relation, and the action of the channel can make it nonzero. Thus independence is not monotonic under postprocessing. Similarly, it might also seem that independence cannot increase for a convex mixture of two observables. However, one can find three observables such that $\mathrm{Ind}(S_{X_1,Y}) = \mathrm{Ind}(S_{X_2,Y}) = 0$ but $\mathrm{Ind}(S_{X,Y}) > 0$ where the observable X is realized by a convex mixture of X_1, X_2 (see Appendix C). Yet for complementarity (see Sec. VIII B) there is no such problem, and we will postulate its monotonicity under postprocessing and nonincreasing under convex mixtures.

We now outline the postulates for other two measures of independence. Let Ind^f be the measure of full independence, and it it nonzero only if for all possible nontrivial marginals of $\mathbf{q}_X(P), \mathbf{q}_Y(P)$ the independence is nonzero. Formally, we require $\mathrm{Ind}^f(S_{X,Y}) = 0$, if there exists *extremal* stochastic maps Λ_1^e, Λ_2^e such that $\Lambda_1^e X \to \Lambda_2^e Y$ or $\Lambda_2^e Y \to \Lambda_1^e X$ [equivalently, $\mathrm{Ind}((\Lambda_1^e, \Lambda_2^e)S_{X,Y}) = 0$], where Λ_1^e, Λ_2^e corresponds to all possible permutations and coarse graining except the trivial one. Apart from this, Ind^f is required to fulfill postulates (23) and (24) of independence.

Let us denote the measure of single-outcome independence by $\mathrm{Ind}^1(S_{X,Y})$. We require $\mathrm{Ind}^1(S_{X,Y}) = 0$ if there exists extremal stochastic maps Λ_1^e, Λ_2^e that belong to a class of coarse grainings resulting in a binary-outcome observable, in which exactly $(n-1)$ outcomes are coarse grained to one outcome, such that $\mathrm{Ind}((\Lambda_1^e, \Lambda_2^e)S_{X,Y}) = 0$. In addition, it should also satisfy the other postulates (23) and (24) of independence.

B. Complementarity

The postulates for complementarity are as follows:

(1) $\mathrm{C}_{X,Y} \geqslant 0$, and $\mathrm{C}_{X,Y} = 0$ if there exists another observable Z in the theory such that $Z \to X$ and $Z \to Y$.

(2) A measure of complementarity cannot increase if instead of X and Y we have only access to statistics of postprocessed observables. Mathematically, this corresponds to

$$\mathrm{C}_{X,Y} \geqslant \mathrm{C}_{\Lambda_1 X, \Lambda_2 Y}, \tag{25}$$

where $\Lambda_{1,2}$ are arbitrary stochastic $n \times n$ matrices. As a consequence, any complementarity measure is invariant under stochastic maps $\Lambda_{1,2}$ whose inverses are also stochastic maps. For example, $\mathrm{C}_{\pi_1 X, \pi_2 Y} = \mathrm{C}_{X,Y}$ for all permutations $\pi_{1,2}$ of an n-element set.

(3) Complementarity cannot increase under a mixture of observables, i.e.,

$$\mathrm{C}_{X,Y} \leqslant \alpha \mathrm{C}_{X_1,Y} + (1-\alpha)\mathrm{C}_{X_2,Y}, \tag{26}$$

where the observable X is realized by the convex mixture of two observables X_1, X_2 with probability distribution $(\alpha, 1-\alpha)$.

Remark. Qualitatively, postulate 3 can be justified by postulate 1. Specifically if observables $X_{1,2}$ are not complementary with Y (i.e., $\mathrm{C}_{X_1,Y} = \mathrm{C}_{X_2,Y} = 0$), then observable X, realized by their convex mixture (with weights α and $1-\alpha$, respectively), is also not complementary with Y. Indeed as a mother observable of X and Y one can take a mixture (with the same weights as above) of mother observables O_1 and O_2 of pairs (X_1, Y) and (X_2, Y). This works because without loss of generality the stochastic maps $O_1 \to (X_1, Y)$, $O_1 \to (X_2, Y)$ can be taken as simply taking marginals.

To see the connection between independence and complementarity recall that the former can be used to define the latter. Concretely, using the prescription from Eq. (9) we obtain that any measure of independence defines that we need the following notions. Now given an independence measure one can obtain the complementarity measure for two extremal observables as follows,

$$\mathrm{C}_{X,Y} = \min_{X':X'\to X} \min_{Y':Y'\to Y} \mathrm{Ind}(S_{X',Y'}), \tag{27}$$

where the infimum is taken over all pairs of observables X', Y' that simulate a pair X, Y. For general observables, we follow the convex-roof extension of the above definition (27). Formally,

$$\mathrm{C}_{X,Y} = \min_{\{\alpha_i, X_i\}} \min_{\{\beta_j, Y_j\}} \sum_{i,j} \alpha_i \beta_j\, \mathrm{C}_{X_i,Y_j}, \tag{28}$$

where the minimum is taken over all possible decompositions of the observables X, Y to the extremal observables, i.e.,

$$\forall P \in \mathcal{P},\ \mathbf{q}_X(P) = \sum_i \alpha_i \mathbf{q}_{X_i}(P),\ \mathbf{q}_Y(P) = \sum_j \beta_j \mathbf{q}_{Y_j}(P). \tag{29}$$

We can express complementarity from a measure of independence in the explicit form as follows:

$$\mathrm{C}_{X,Y} = \min_{\{\alpha_i, X_i\}} \min_{\{\beta_j, Y_j\}} \sum_{i,j} \alpha_i \beta_j \min_{X_i':X_i'\to X_i} \min_{Y_j':Y_j'\to Y_j} \mathrm{Ind}(S_{X_i',Y_j'}). \tag{30}$$

Thus, for clean and extremal observables, C = Ind. Note that while independence was not required to be monotonic under stochastic maps, due to the above definition of complementarity it will be natural to require such monotonicity.

Remark. As we have said in Sec. IV, the simplest theory for which complementarity is not equal to independence is the already mentioned two bits with three observables: X for the first bit, Y for the second bit, and the third observable Z with four outcomes that measures the value of both bits. The two observables X and Y are clearly independent for any possible measure, while they both come from C by postprocessing, so that they are not clean, and complementarity vanishes.

Similarly, one can set the postulates of the measures of full complementarity and single-outcome complementarity. We denote the measures by $\mathrm{C}^f_{X,Y}$ and $\mathrm{C}^1_{X,Y}$, respectively. $\mathrm{C}^f(S_{X,Y}) = 0$ if there exist *extremal* stochastic maps Λ_1^e, Λ_2^e such that $\mathrm{C}((\Lambda_1^e, \Lambda_2^e)S_{X,Y}) = 0$, where Λ_1^e, Λ_2^e corresponds to all possible permutations and coarse graining except the trivial one, while $\mathrm{C}^1_{X,Y} = 0$ if there exist extremal stochastic maps Λ_1^e, Λ_2^e that belong to a class of coarse grainings resulting in a binary-outcome observable, in which exactly $(n-1)$ outcomes are coarse grained to one outcome, such that $\mathrm{C}((\Lambda_1^e, \Lambda_2^e)S_{X,Y}) = 0$. Additionally, both the measures should satisfy the postulates of nonincreasing under postprocessing (25).

IX. MEASURES OF UNCERTAINTY AND INDEPENDENCE

In this section, we propose some measures of uncertainty and independence.

A. Complementarity and uncertainty measures based on random access code

We propose a measure of independence based on a communication task known as random access code [51]. This task involves two devices, preparation and measurement, possessed by Alice and Bob, respectively. In each round of the task, Alice receives a two-bit input $a = (a_1 a_2) \in \{1, \dots, d\}^2$, prepares a d-dimensional system, say P_a, and sends to Bob. Bob receives the communicated system from Alice and measures an observable depending on his obtained input $b \in \{1, 2\}$. He wants to guess a_b. Let us denote the probability of giving the correct answer for input a, b as $p(a_b|a, b)$. A figure of merit of such communication task can be any reasonable function of these probabilities, $\mathcal{F}\{p(a_b|a, b)\}$. For instance, it could be the average success probability of guessing a_b,

$$p_s = \frac{1}{2d^2} \sum_{a,b} p(a_b|a, b), \tag{31}$$

where the inputs are uniformly distributed.

In the most common version of the above task, Bob is free to choose the optimal observables that would maximize the probability of success. Here, to connect the task with complementarity, we will fix Bob's observables to be one of two observables X and Y. For convenience let us denote $X = X_1$ and $Y = X_2$. Now, for input a, b, Bob obtains a statistics $\mathbf{q}_{X_b}(P_a)$ where X_b denotes the d-outcome observable measured on P_a. He can apply some postprocessing after the measurement, and thus the obtained probability for the correct answer is

$$p(a_b|a,b) = \tilde{\mathbf{q}}_{X_b}(a_b|P_a), \text{ where } \tilde{\mathbf{q}}_{X_b}(P_a) = \Lambda_b \mathbf{q}_{X_b}(P_a). \quad (32)$$

Now given any theory and the two observables X_1, X_2, the relevant quantity $p_s = \mathcal{F}\{p(a_b|a,b)\}$ is maximized over all possible P_a, Λ_b. Let the measure of independence of these two observables be as follows,

$$\mathrm{Ind}(S_{X_1,X_2}) = \frac{p_s(X_1,X_2) - \max\,(p_s(X_1), p_s(X_2))}{1 - \max\,(p_s(X_1), p_s(X_2))}, \quad (33)$$

where $p_s(X_1, X_2)$ denotes the optimal value of the figure of merit when Bob has access to two observables X_1, X_2 and $p_s(X_1)$ denotes the same when Bob has access to only X_1. Note that $\mathrm{Ind}(S_{X_1,X_2})$ is normalized; i.e., it takes value within the range [0,1].

One can readily check that the measure (33) satisfies the postulates of independence. Since Bob is allowed to apply arbitrarily stochastic map Λ_b, $p_s(X_1, X_2)$ is eventually equal to $p_s(X_1)$ [or $p_s(X_2)$] if $X_1 \to X_2$ (or $X_2 \to X_1$). Due to the same reason, it is invariant under permutation. Further, as $\mathcal{F}\{p(a_b|a,b)\}$ is maximized over all possible preparations P_a, it is monotonic under inclusion.

This measure relates independence to efficacy of an operational task. However, it is not a measure of full independence. In future, one may look for a similar operational task that quantifies full independence.

One can define a measure of uncertainty based on the same communication task. In this situation, Bob is allowed to apply only a doubly stochastic map on the observed statistics after measurement. The measure of uncertainty for a preparation P is considered to be the converse of the maximum success probability of guessing a_b over all possible inputs a,

$$\begin{aligned}\mathrm{U}(\mathbf{q}_{X_1}(P), \mathbf{q}_{X_2}(P)) &= 1 - \max_a \frac{1}{2} \sum_b p(a_b|a,b) \\ &= 1 - \max_a \frac{1}{2}(q_{X_1}(a_1|P) + q_{X_2}(a_2|P)).\end{aligned} \quad (34)$$

Subsequently, following (16), the uncertainty of the statistics set

$$\mathrm{U}(S_{X_1,X_2}) = 1 - \max_{P_a \in \mathcal{P}} \frac{1}{2} \sum_b q_{X_b}(a_b|P_a). \quad (35)$$

By the definition the above measure (34) is zero if and only if the distribution of $\mathbf{q}(P)$ is deterministic and cannot decrease under the doubly stochastic map. The measure of uncertainty can be rewritten as $\min_{P_a\in\mathcal{P}} (1 - \frac{1}{2}\sum_b q_{X_b}(a_b|P_a))$. Since $1 - \frac{1}{2}\sum_b q_{X_b}(a_b|P_a)$ is linear with respect to a convex mixture of two preparations and the minimum function of two linear functions is concave, it satisfies (14). It can also be readily checked that $\frac{1}{2}\sum_b q_{X_b}(a_b|P_a)$ cannot decrease under coarse graining of observables, and therefore it satisfies monotonicity under coarse graining (13). To see that the measure also satisfies convexity (15), we express the uncertainty measure (34) between X_2 and a convex mixture of two observables X_1, X_1' with probability distribution $(\alpha, 1-\alpha)$ in the following way:

$$\begin{aligned}&1 - \max_a \tfrac{1}{2}[\alpha q_{X_1}(a_1|P) + (1-\alpha) q_{X_1'}(a_1|P) + q_{X_2}(a_2|P)] \\ &\geqslant \alpha\{1 - \max_a \tfrac{1}{2}[q_{X_1}(a_1|P) + q_{X_2}(a_2|P)]\} \\ &\quad + (1-\alpha)\{1 - \max_a \tfrac{1}{2}[q_{X_1'}(a_1|P) + q_{X_2}(a_2|P)]\} \\ &= \alpha \mathrm{U}(\mathbf{q}_{X_1}(P), \mathbf{q}_{X_2}(P)) + (1-\alpha)\mathrm{U}(\mathbf{q}_{X_1'}(P), \mathbf{q}_{X_2}(P)).\end{aligned} \quad (36)$$

Let us remark that a variant of the obtained measure of uncertainty was considered, e.g., in [14]. Here we have pointed out its operational origin (by connecting it to the random access code), as well as shown that it satisfies the postulates. Similarly, we define the measure of information exclusion as the converse of the average success probability of guessing a_b restricted to those inputs when $a_1 = a_2$,

$$\begin{aligned}\mathrm{E}(S_{X_1,X_2}) &= 1 - \frac{1}{2d} \sum_{b,a|a_1=a_2} p(a_b|P_a) \\ &= 1 - \frac{1}{2d} \max_{P_a\in\mathcal{P}} \sum_{b,a|a_1=a_2} \tilde{q}_{X_b}(a_b|P_a),\end{aligned} \quad (37)$$

taking into account $\tilde{\mathbf{q}}_{X_b}(P_a) = \pi \mathbf{q}_{X_b}(P_a)$.

Example: Quantum theory. To provide a complete example in quantum theory, we take the figure of merit as the average success probability (31). It has been shown that the optimal value for a classical system [52]

$$p_s(X_1) = \frac{1}{2} + \frac{1}{2d}. \quad (38)$$

For the two quantum projective measurements corresponding to the bases $X_1 = \{|i\rangle\}_{i=1}^d$ and $X_2 = \{|\psi\rangle_j\}_{j=1}^d$ accessed by Bob, the average success probability (31)

$$p_s(X_1, X_2) = \frac{1}{2} + \frac{1}{2d^2} \sum_{a_1,a_2} |\langle a_1|\psi_{a_2}\rangle|. \quad (39)$$

The proof of this fact is given in Appendix D. The left-hand side of (39) is strictly greater than $p_s(X_1)$ (38) for any two distinct quantum observables since

$$\sum_{a_1,a_2} |\langle a_1|\psi_{a_2}\rangle| > \sum_{a_1,a_2} |\langle a_1|\psi_{a_2}\rangle|^2 = d. \quad (40)$$

The optimal quantum value of $p_s = \frac{1}{2} + \frac{1}{2\sqrt{d}}$, which corresponds to two mutually unbiased bases [53].

Hence, the independence measure (33) based on the random access code for two d-dimensional quantum observables is given by

$$\mathrm{Ind}(S_{X_1,X_2}) = \frac{1}{d-1}\left(\frac{1}{d}\sum_{a_1,a_2} |\langle a_1|\psi_{a_2}\rangle| - 1\right). \quad (41)$$

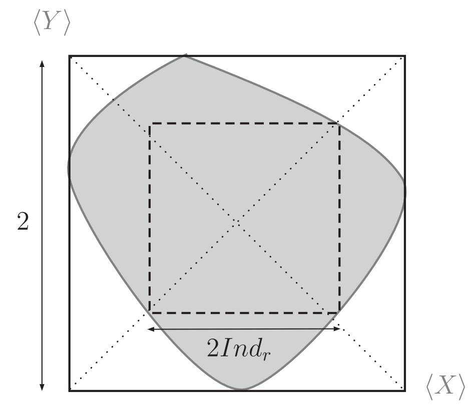

FIG. 8. The rescaling measure of independence (Ind_r) of the statistics set for binary-outcome observables X, Y. The statistics set is presented in gray.

Further, invoking (D1) one obtains the uncertainty measure (35)

$$\mathrm{U}(S_{X_1,X_2}) = \frac{1}{2}\big(1 - \max_{a_1,a_2} \big|\langle a_1|\psi_{a_2}\rangle\big|\big) \tag{42}$$

and the information exclusion measure (37)

$$\mathrm{E}(S_{X_1,X_2}) = \frac{1}{2}\left(1 - \frac{1}{d}\max_{\pi} \sum_{i=\pi(j)} |\langle a_i|\psi_{a_{\pi(j)}}\rangle|\right), \tag{43}$$

where π is d-element permutation.

B. Rescaling and volume of the probability space

We shall now define measure of independence by means of rescalings of the statistics set $S_{X,Y}$.

Definition 14. Independence is given by maximal $r \in [0, 1]$ such that $rS + x \subset S_{X,Y}$. That is, r is a maximum rescaling factor of the full set S such that the rescaled set $rS + x$ is contained in S after shifting along some vector x. We denote it by Ind_r (see Fig. 8).

It is clear from the definition that Ind_r is invariant under permutation (23) and monotonic under inclusion (24). We know that the dimension of S is $2(d-1)$. If $\Lambda_1^e X \to \Lambda_2^e Y$ for some extremal stochastic maps $\Lambda_{1,2}^e$, then the number of independent variables to specify $\mathbf{q}(P)$ is less than $2(d-1)$. It follows that the dimension of the statistics set $S_{X,Y}$ is strictly less than $2(d-1)$; thereby $\mathrm{Ind}_r = 0$. Thus, Ind_r is a good measure of full independence. For instance, S being a square (i.e., two binary observables) the full set s-bit and classical c-bit have complementarities 1, 0, respectively. In the quantum case, consider qubit observables $Z = \sigma_z$, $X = \mathbf{n}\cdot\sigma$ with $n_y = 0$ and $n_x^2 + n_z^2 = 1$. The boundary of the statistics set of possible pairs of averages $(\langle\psi|Z|\psi\rangle, \langle\psi|X|\psi\rangle)$ is given by

$$\frac{(x+z)^2}{2a^2} + \frac{(x-z)^2}{2b^2} = 1, \text{ with } a = \frac{n_x}{\sqrt{1-n_z}}, b = \frac{n_x}{\sqrt{1+n_z}}. \tag{44}$$

This is shown in Fig. 9. The parameters a and b are the major and minor semiaxes of the ellipse, respectively. Thus, the diagonal of the largest square inside the body is $2b$.

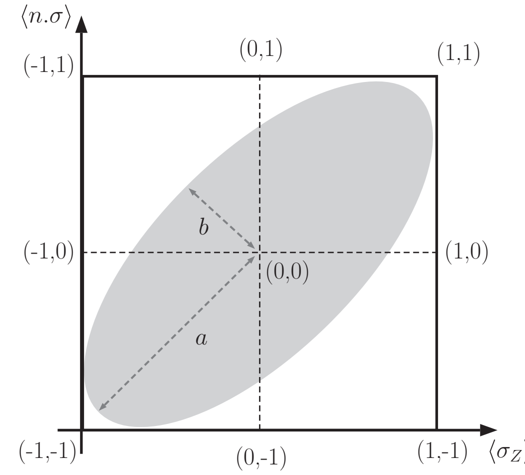

FIG. 9. The statistics set of two quantum observables σ_z and $n_x\sigma_x + n_z\sigma_z$ is presented in gray. The semimajor and semiminor axes are denoted by a and b, respectively.

Subsequently, a simple calculation leads to

$$\mathrm{Ind}_r = \frac{\sqrt{2}b}{2} = \frac{n_x}{\sqrt{2(1+n_z)}}. \tag{45}$$

Note that with this definition, the q-bit does not have the maximal possible complementarity of the s-bit.

Following the same arguments, one can see that the volume of $S_{X,Y}$ is also a measure of full independence. For s-bit and q-bit observables (Z, X) and the c-bit the volume of $S_{X,Y}$ is 4, $\pi ab = \pi n_x$, and 0, respectively.

Here, it can be noted that the uncertainty (U) (35) proposed in Sec. IX A is zero for the s-bit and c-bit, while for the quantum observables in Fig. 9,

$$\begin{aligned}\mathrm{U} &= 1 - \max_{P_a\in\mathcal{P}} \frac{1}{2}\sum_b q_{X_b}(a_b|P_a)\\ &= 1 - \frac{a}{\sqrt{2}} = 1 - \frac{n_x}{\sqrt{2(1-n_z)}}.\end{aligned} \tag{46}$$

C. Complementarity measures based on preimage

In this section, we shall propose just a scheme of building various measures of independence from a class of functions defined on joint distributions. Namely, we will require from such a function that it vanish on distributions of the form $p(i,j) = p(i,j)\delta_{ij}$. We shall slightly abuse notation by naming such functions also "independence" (now not independence of a pair of observables, but independence of joint distribution). An example of the independence measure is the so-called *variation of information*:

$$VI(p_{XY}) = H(X|Y) + H(Y|X), \tag{47}$$

where $H(\cdot)$ is the entropy. To define independence on pairs of observables from that defined on a joint distribution we proceed as follows. Fix some set S_{pre} to be a convex set of joint distributions, whose marginals give rise to $S_{X,Y}$. Let us fix two channels Λ_1 and Λ_2 acting on the outputs of observables X and Y, respectively. We now consider a set $S_{pre}(X, Y, \Lambda_1, \Lambda_2)$ (in short S_{pre}) of joint distributions which after applying local

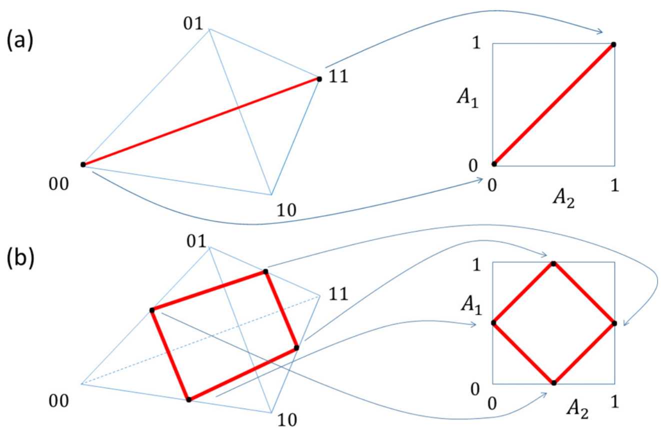

FIG. 10. Examples of a classical bit and a "diamond." (a) Independence of two identical observables is zero, since it is obtained as an image of perfectly correlated probability distributions. (b) Independence of observables for which the statistics constitutes a diamond is equal to 1. Preimage is a square that contains the center of the product of the simplices, which has independence 1.

processing $\Lambda_1 \otimes \Lambda_2$, where Λ_i are channels, gives rise to $S_{X,Y}$ via marginals. In other words, each element of $S_{X,Y}$ is a pair of marginals of some distribution from S_{pre} subjected to $\Lambda_1 \otimes \Lambda_2$, and vice versa, if we apply the $\Lambda_1 \otimes \Lambda_2$ channel to each joint distribution from S_{pre}, the pair of marginals of the obtained distribution belongs to $S_{X,Y}$.

The independence measure is now defined as

$$\mathrm{Ind}(X, Y) = \min_{S_{pre}} \max_{p \in S_{pre}} \mathrm{Ind}(p), \tag{48}$$

where the minimum is taken over all convex sets S_{pre} of distributions, such that there exist channels Λ_1 and Λ_2 for which S_{pre} gives rise to $S_{X,Y}$, as described above.

Let us see that the measure satisfies the postulates for independence. Suppose that one observable is a processed version of the other, i.e., can be obtained from the other via some channel Λ. Then we can take the preimage to be the set of perfectly correlated distributions, with the choice $\Lambda_1 \otimes \Lambda_2 = I \otimes \Lambda$. Hence all the distributions from the preimage have vanishing independence, so that the measure vanishes. By definition, if we enlarge the set $S_{X,Y}$, the measure can only increase, as the preimage cannot decrease. Thus we obtain that the second postulate is satisfied too.

We illustrate the concept of the above measure by means of two examples: the classical bit, in Fig. 10(a), and the "diamond" bit, in Fig. 10(b), where we take the variation of information as the independence measure of joint distributions.

For the classical bit (two identical observables) the set $S_{X,Y}$ can be obtained as an image of an edge of the tetrahedron, which allows only for perfectly correlated distributions; hence the measure vanishes.

Let us argue that the set depicted in Fig. 10(b) is the only possible preimage. Note first that corners of the diamond are the following pairs of distributions (we use quantum notation just for brevity):

$$(I/2, |0\rangle\langle 0|), (I/2, |1\rangle\langle 1|), (|0\rangle\langle 0|, I/2), (|1\rangle\langle 1|, I/2). \tag{49}$$

Since always one of the distributions in the pair is pure, the only joint distributions that return these pairs via marginals are product. Let us argue that for any fixed pair of channels $\Lambda_1 \otimes \Lambda_2$, the distributions that can give rise through these channels to product distributions must be product too. To this end, note that if we start with a correlated distribution, and act with a product channel, the output distribution is product if and only if at least one of the channels is "information killing"; i.e., it produces a single state for all input states. Clearly none of our channels can be like that, because sometimes we need to produce $I/2$ and sometimes $|0\rangle\langle 0|$ or $|1\rangle\langle 1|$. Thus, the initial joint distributions must be product.

The channel Λ_1 has just to send two of the distributions to $I/2$, one to $|0\rangle\langle 0|$, and one to $|1\rangle\langle 1|$ (the same about channel Λ_2). Suppose that the distribution sent to $|0\rangle\langle 0|$ is neither $|0\rangle\langle 0|$ nor $|1\rangle\langle 1|$. Then one directly checks that the channel sends all the states to $|0\rangle\langle 0|$, which cannot be so (as we want also to get $|1\rangle\langle 1|$ and $I/2$ for some input states). Thus the input must be either $|0\rangle\langle 0|$ or $|1\rangle\langle 1|$. Suppose it is $|0\rangle\langle 0|$. Then one finds that the channels is of the form

$$\begin{bmatrix} 1 & q \\ 0 & 1-q \end{bmatrix}. \tag{50}$$

Now this channel must produce $|1\rangle\langle 1|$ out of some state. One finds then that the channel must be identity. If the input is $|1\rangle\langle 1|$ we obtain that the channel is a flip. Similarly Λ_2 is either identity of a flip. Thus the preimage of the four corners of the diamond are the products

$$I/2 \otimes |0\rangle\langle 0|,\ I/2 \otimes |1\rangle\langle 1|,\ |0\rangle\langle 0| \otimes I/2,\ |1\rangle\langle 1| \otimes I/2. \tag{51}$$

Hence the preimage, since it is a convex set by definition, contains $I/2 \otimes I/2$ as an equal mixture of the above distributions. We conclude that the measure of independence is equal to 1.

X. PREPARATION UNCERTAINTY RELATION

As proposed in Sec. III from measures of uncertainty and complementarity, one can build uncertainty relations of the form

$$\mathrm{U}(S_{X,Y}) \geqslant f^{\uparrow}(\mathrm{C}(S_{X,Y})), \tag{52}$$

where $f^{\uparrow}$ is a nondecreasing function whose range is nonnegative.

We first note that such uncertainty principle is not satisfied in all theories. For example, the square bit, whose statistics set is the whole square, cannot satisfy the above uncertainty relations for any measures of complementarity. Indeed from postulates it follows that if $S_{X,Y}$ is the whole square, then there is no uncertainty of any kind, as it contains all corners. Also, complementarity, by monotonicity under inclusion, must be maximally possible. Therefore, any complementarity measure (apart from the trivial one that is zero for all possible sets) will be nonzero.

A. PUR from random access codes

We derive here the PUR constructed out of measures of uncertainty and complementarity in terms of random access codes from Sec. IX A. This PUR is actually the exclusion principle of a form similar to that of [41].

Fact 1. In quantum mechanics the following PUR holds for arbitrary two observables X and Y with d outcomes, with one-dimensional eigenprojectors:

$$\mathrm{E}(S_{X,Y}) \geqslant \frac{(\mathrm{C}_{X,Y})^2}{4d}, \tag{53}$$

where E is the measure of exclusion of (43) and $\mathrm{Ind}(S_{X,Y}) = \mathrm{C}_{X,Y}$ is the measure of independence of (41).

Of course, since the considered observables are clean and extremal Ind is the same as complementarity.

B. PUR from rescaling

Here we consider the rescaling measures of complementarity ($\mathrm{Ind}_r = \mathrm{C}_r$ for clean and extremal observables) and uncertainty U mentioned in Sec. IX B to provide an example of the PUR between binary observables.

Fact 2. Two quantum binary observables $Z = \sigma_z$, $X = \mathbf{n}\cdot\sigma$ with $n_y = 0$ and $n_x^2 + n_z^2 = 1$ satisfy the following PUR, which is even in a form of equality:

$$\mathrm{C}_r^2 + (1-\mathrm{U})^2 = 1. \tag{54}$$

Proof. First one can express Ind_r and U in (45) and (46) in terms of only n_z by substituting $n_x = \sqrt{1-n_z^2}$. Further, by equalizing n_z as a function of C_r and U, one obtains the above PUR with equality.

C. Reverse PUR from rescaling

In Sec. III E we introduced the concept of the reverse uncertainty relation. As said there, unlike the uncertainty relation, which may or may not hold in a given theory, the reverse one is expected to hold almost by definition in any theory. Here we present such a relation in the case of binary outcomes, for the uncertainty based on rescaling.

Fact 3. In any theory, for any two binary sharp, clean, and extremal observables, the following reverse PUR holds:

$$2\mathrm{C}_r \geqslant \mathrm{U}. \tag{55}$$

The proof is given in Appendix F.

D. Uncertainty relation from physical principles

Now, we shall show how the information-theoretic principle, namely the *information contents principle* [47], a single-system version of *information causality* [46], imposes the PUR on the physical theories. Likewise, one can postulate the PUR or obtain the PUR from other principles which should be obeyed by any physical theories.

Let us recall the communication task random access code presented before. We assume the inputs a, b, given to Alice and Bob, are uniformly distributed and uncorrelated, i.e., $\forall a, b,\ p(a,b) = p(a)p(b),\ p(a) = 1/4,\ p(b) = 1/2$. We denote the classical output of Bob by C_b for his input b. The information causality provides a bound on the correlations as follows,

$$\begin{aligned} &I(C_1 : X) + I(C_2 : Y) - I(C_1 : C_2) \leqslant 1 \\ &\quad \Rightarrow H(X) - H(C_1X) + H(Y) - H(C_2Y) + H(C_1C_2) \leqslant 1. \end{aligned} \tag{56}$$

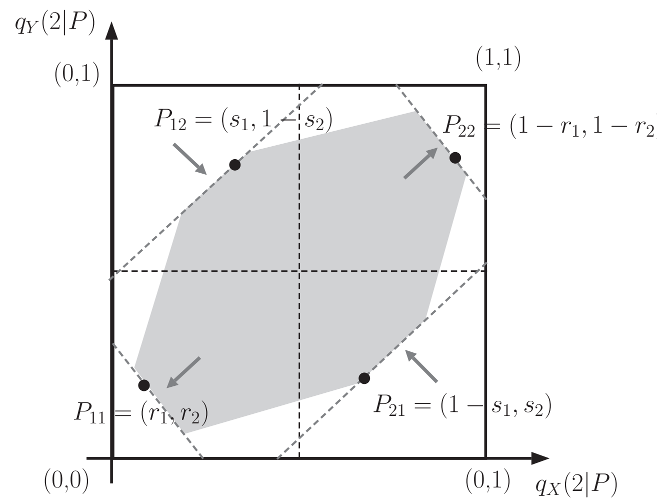

FIG. 11. The statistics set $S_{X,Y}$ possesses the symmetry under the reflection of the diagonal of the square. State of the system $P_{a_1a_2}$ is described by the pair of probabilities $(q_X(2|P), q_Y(2|P))$.

Since we deal with two binary-outcome measurements, the statistics set $S_{X,Y}$ can be conveniently presented by the pair of probabilities $(q_X(2|P), q_Y(2|P))$ as shown in Fig. 11. For the sake of simplicity, we consider a class of theories in which the statistics set $S_{X,Y}$ possesses symmetry under permutation of outcome; i.e., for all $q_X(P)$ there exists another preparation P' such that $q_Y(P') = q_X(P)$ and vice versa. In other words, $S_{X,Y}$ is symmetric with respect to the diagonal of the square. Due to the symmetry of the statistics set in Fig. 11, for a preparation with statistics $(q_X(2|P), q_Y(2|P)) = (r_1, r_2)$, we know there is another preparation with $(q_X(2|P'), q_Y(2|P')) = (1-r_1, 1-r_2)$. Accordingly, we obtain the probability distribution for C_1X and C_2Y:

	$X=1$	$X=2$
$C_1=1$	$\frac{1}{2}\left(\frac{r_1}{2}+\frac{s_1}{2}\right)$	$\frac{1}{2}\left(1-\frac{r_1}{2}-\frac{s_1}{2}\right)$
$C_1=2$	$\frac{1}{2}\left(1-\frac{r_1}{2}-\frac{s_1}{2}\right)$	$\frac{1}{2}\left(\frac{r_1}{2}+\frac{s_1}{2}\right)$
	$Y=1$	$Y=2$
$C_2=1$	$\frac{1}{2}\left(\frac{r_2}{2}+\frac{s_2}{2}\right)$	$\frac{1}{2}\left(1-\frac{r_2}{2}-\frac{s_2}{2}\right)$
$C_2=2$	$\frac{1}{2}\left(1-\frac{r_2}{2}-\frac{s_2}{2}\right)$	$\frac{1}{2}\left(\frac{r_2}{2}+\frac{s_2}{2}\right)$

Thus,

$$\begin{aligned} &H(X) = H(Y) = \frac{1}{2}H(C_1C_2) = 1, \\ &H(C_1X) = h\left(\frac{r_1}{2}+\frac{s_1}{2}\right)+1, H(C_2Y) = h\left(\frac{r_2}{2}+\frac{s_2}{2}\right)+1, \end{aligned} \tag{57}$$

where $h(p) = -p\log_2(p) - (1-p)\log_2(1-p)$. Substituting these expressions in the ICP (56) we obtain the following relation,

$$h\left(\frac{r_1}{2}+\frac{s_1}{2}\right) + h\left(\frac{r_2}{2}+\frac{s_2}{2}\right) \geqslant 1. \tag{58}$$

Notably, the above relation coincides with the Maassen-Uffink uncertainty relation [36] of σ_x, σ_z. By taking values of the parameters $r_{1,2}, s_{1,2}$ in the small interval, one can see that the above relation (58) is satisfied if

$$r_1 + r_2 + s_1 + s_2 \geqslant 0.44. \quad (59)$$

This relation is valid for any two given preparations P_{11}, P_{12}. Thanks to the symmetry, there exists a preparation on the diagonal of the square that corresponds to the minimum uncertainty of all possible preparations, i.e., the uncertainty of $S_{X,Y}$. Again, exploiting the symmetry one knows that the origin of the largest square fit inside $S_{X,Y}$ is the center of the square. Therefore, for the symmetric statistics set,

$$\mathrm{U} = 2\min(r, s),\ \mathrm{C}_r = 1 - 2\max(r, s), \quad (60)$$

where $r_1 = r_2 = r, s_1 = s_2 = s$. Subsequently, it follows from (59) that $\mathrm{C}_r - \mathrm{U} \leqslant 0.56$ which captures the PUR. Namely, the last formula says that for strong enough complementarity uncertainty must appear.

E. Tsirelson bound from uncertainty principle and nonsignaling

Here, we discuss how the uncertainty principle in a theory sets a restriction on the nonlocality of that theory. We concentrate on the simplest scenario of nonlocality where two spatially separated parties, Alice and Bob, perform one of the two binary-outcome measurements $A_{1,2}, B_{1,2} \in \{+, -\}$ on their respective subsystems of a bipartite system. The witness based on the measurement statistics of nonlocality is taken to be the violation of the well-known Clauser-Horne-Shimony-Holt (CHSH) local-realist inequality [54],

$$\mathcal{I} = \langle A_1B_1\rangle + \langle A_1B_2\rangle + \langle A_2B_1\rangle - \langle A_2B_2\rangle \leqslant 2. \quad (61)$$

Without loss of generality, we can say that Bob's measurement statistics of the observables B_1, B_2 on his system are $\mathbf{q}_{B_1}(P), \mathbf{q}_{B_2}(P)$ for some P when Alice does not perform any measurement. As a result of sharing correlated systems, depending on Alice's measurement choice and outcome the preparation on Bob's side might be different. In other words, Alice's measurement steers a different preparation on Bob's subsystem. Let us denote Bob's preparation as P_{A_1+} if Alice measures A_1 and obtains $+$ outcome on her subsystem and so on. This phenomenon is called "steering" [14] [55]. However, we do not impose any restriction on steering, except the no-signaling principle which should be satisfied by any physical theory. The "no-signaling" principle is a direct consequence of relativistic causation, which says that Alice cannot send any information to Bob instantaneously. That is, the measurement statistics on Bob's subsystem is independent of Alice's measurement choice and vice versa. Formally, $\forall i \in \{1, 2\}$,

$$\begin{aligned}\mathbf{q}_{B_i}(P) &= q_{A_1}(+|\tilde{P})\mathbf{q}_{B_i}(P_{A_1+}) + q_{A_1}(-|\tilde{P})\mathbf{q}_{B_i}(P_{A_1-})\\ &= q_{A_2}(+|\tilde{P})\mathbf{q}_{B_i}(P_{A_2+}) + q_{A_2}(-|\tilde{P})\mathbf{q}_{B_i}(P_{A_2-}),\end{aligned} \quad (62)$$

where $\tilde{P}$ denotes Alice's initial preparation. For simplicity, we denote

$$\begin{aligned}&q_{A_1}(+|\tilde{P}) = t_1,\ q_{A_2}(+|\tilde{P}) = t_2,\\ &q_{B_1}(+|P_{A_1+}) = 1 - r_1,\ q_{B_2}(+|P_{A_1+}) = 1 - r_2,\\ &q_{B_1}(+|P_{A_1-}) = r_1',\ q_{B_2}(+|P_{A_1-}) = r_2',\\ &q_{B_1}(+|P_{A_2+}) = 1 - s_1,\ q_{B_1}(+|P_{A_2+}) = s_2,\\ &q_{B_2}(+|P_{A_2-}) = s_1',\ q_{B_2}(+|P_{A_2-}) = 1 - s_2',\end{aligned} \quad (63)$$

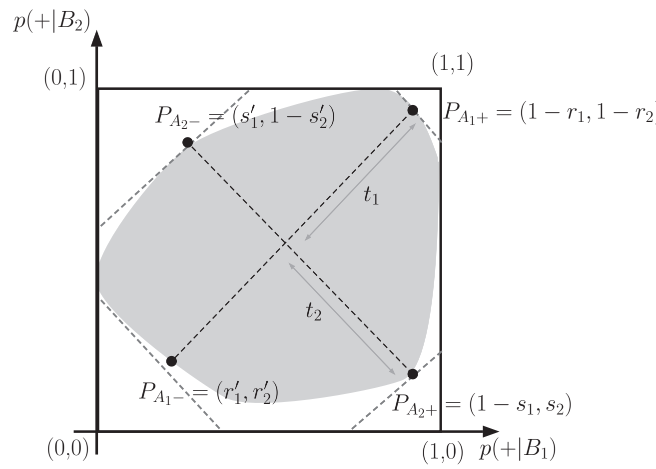

FIG. 12. An arbitrary statistics set $S_{X,Y}$ for two observables $B_{1,2}$ of Bob's system. The four different preparations $P_{A_i\pm}$ depending on Alice's measurement choice and outcome are presented by their coordinates. The four preparations should satisfy the no-signaling conditions (65).

as shown in Fig. 12. Subsequently, the CHSH term is expressed as follows,

$$\begin{aligned}\mathcal{I} &= q_{A_1}(+|\tilde{P})[2 - 2q_{B_1}(-|P_{A_1+}) - 2q_{B_2}(-|P_{A_1+})]\\ &\quad + q_{A_1}(-|\tilde{P})[2 - 2q_{B_1}(-|P_{A_1-}) - 2q_{B_2}(-|P_{A_1-})]\\ &\quad + q_{A_2}(+|\tilde{P})[2 - 2q_{B_1}(-|P_{A_2+}) - 2q_{B_2}(+|P_{A_2+})]\\ &\quad + q_{A_2}(-|\tilde{P})[2 - 2q_{B_1}(-|P_{A_2-}) - 2q_{B_2}(+|P_{A_2-})]\\ &= 4 - 2[t_1(r_1 + r_2) + (1 - t_1)(r_1' + r_2') + t_2(s_1 + s_2)\\ &\quad + (1 - t_2)(s_1' + s_2')],\end{aligned} \quad (64)$$

while the no-signaling conditions simplify to

$$\begin{aligned}t_1(1 - r_1) + (1 - t_1)r_1' &= t_2(1 - s_1) + (1 - t_2)s_1',\\ t_1(1 - r_2) + (1 - t_1)r_2' &= t_2s_2 + (1 - t_2)(1 - s_2').\end{aligned} \quad (65)$$

Thus, we seek to maximize the right-hand side of (64) under the nonlinear constraints (65). Intuitively, it can be seen that the PUR prevents the CHSH value from being the maximum. There are only a few possibilities for $\mathcal{I} = 4$. In one case, the statistics set allows the four corners of the square, i.e., $r_1 + r_2 = r_1' + r_2' = s_1 + s_2 = s_1' + s_2' = 0$, which contradicts the notion of the PUR. On the other, one of the terms $r_1 + r_2$ or $r_1' + r_2'$, say $r_1 + r_2$, and one of terms $s_1 + s_2$ or $s_1' + s_2'$, say $s_1 + s_2$, is zero and accordingly t_1, t_2 both have to be 1. Such value assignment of these variables contradicts the no-signaling principle (65).

If we assume $S_{X,Y}$ to be symmetric with respect to the diagonal of the square (as shown in Fig. 11), then it is easier to relate the CHSH term (64) to the uncertainty principle. Consider P_{A_1+}, P_{A_2+} to be the closest points to the corners (1,1) and (1,0), respectively. By symmetry, we know there exists another two closest points to other two corners, such that $r_1 = r_1', r_2 = r_2', s_1 = s_1', s_2 = s_2'$. Therefore, $\mathcal{I} \leqslant 4 - 2(r_1 + r_2 + s_1 + s_2)$. In fact, this inequality is tight, due to the fact that this value is achieved when the no-signaling conditions (65) are satisfied for $t_1 = t_2 = 1/2$. Further, we recall the expression of U, C_r from (60) in terms of r_1, r_2, s_1, s_2, and reexpress the CHSH term as

$$\mathcal{I} = 2 + 2(\mathrm{C}_r - \mathrm{U}). \tag{66}$$

Clearly, the uncertainty principle, which is in the form (52), restricts the value of $\mathcal{I}$. In quantum theory, the exact form of the PUR is given in (54). Thus, the maximum value of the right-hand side of (66) is obtained from the Tsirelson bound, i.e., $2\sqrt{2}$, when $\mathrm{C}_r = 1 - \mathrm{U} = 1/\sqrt{2}$ satisfying (54).

XI. OPEN PROBLEMS

The major open problem is whether there exist theories in which two clean and extremal observables can be very well approximated by some other observable. For such hypothetical theories, complementarity of observables cannot be any longer read out from behavior of the statistics set. It would be also interesting to define a smoothed version of complementarity, given by the minimum of independence over observables that reproduce the given observables up to ϵ in some suitable distance. One can then investigate how the statistics set changes with ϵ. Another interesting problem is to explore the relation between the concepts of complementarity and contextuality [24], as the latter also reflects somewhat the notion of complementarity. There is also a question of how the approach presented in this paper is related to the operational approach to wave particle duality of Ref. [56].

There are many other questions, including the following ones:

(i) Generalize the geometric approach to continuous variables (i.e., to position and momentum observables).

(ii) Prove that the uncertainty relation implies the Tsirelson bound without symmetry assumptions.

(iii) Relate the information contents principle to the uncertainty relation for larger dimensions, and again, without symmetry assumptions.

(iv) Compute independence based on variation of information for qubit observables, and find the uncertainty relation with a properly chosen uncertainty measure (it seems that in this case entropy is the suitable one, or mutual information as an exclusion measure in higher dimensions).

(v) Make tighter the exclusion principle based on random access codes.

Finally, we focused exclusively on two observables, but one can readily extend the definitions and concepts to more observables and explore the subject in this more general setting.

Note added. Recently we became aware of the paper [57], which derived the Tsirelson bound for the CHSH inequality from restrictions on the complementarity present in quantum theory. However, the quantitative notion of complementarity used in that work differs from that considered by us.

ACKNOWLEDGMENTS

We thank Karol Horodecki for interesting and stimulating discussions. D.S. is supported by the National Science Centre, Poland, Grants No. 2016/23/N/ST2/02817 and No. 2014/14/E/ST2/00020, and FNP grant First TEAM (Grant No. First TEAM/2017-4/31). M.O. acknowledges the support of the Homing program of the Foundation for Polish Science cofinanced by the European Union under the European Regional Development Fund. L.C., M.H., and R.H. are supported by the John Templeton Foundation through Grant No. 56033. M.H. and R.H. are also supported by the National Science Centre, Poland, grant OPUS 9, Grant No. 2015/17/B/ST2/01945.

APPENDICES

In the appendices we present proofs of technical results that were omitted from the main text.

APPENDIX A: PROOF OF QUALITATIVE UNCERTAINTY RELATIONS

Lemma 1. In quantum mechanics, for quantum measurements with one-dimensional projectors the following statements hold:

(i) Complementarity implies information exclusion.

(ii) Single-outcome complementarity implies uncertainty.

(iii) Full complementarity implies uncertainty.

Proof. We prove each implication individually.

(i) Note first that two rank-1 projective measurements X and Y are not jointly measurable if and only if they do not commute. In other words some projector P_i of X and some projector Q_j of Y do not commute (see, e.g., [43] for the proof of this statement). Now suppose, by contraposition, that there is no exclusion for X and Y. This means that the d states have distinct deterministic outcomes for observables X and Y. Hence, the states are distinct eigenstates of both observables. Therefore, X and Y commute, hence they are not complementary.

(ii) Again by contraposition, suppose that there is no uncertainty. This means that the observables share a common eigenvector. Consider coarse graining for both observables: this vector versus the complement. Clearly the new binary observables are the same, hence do not exhibit complementarity. Hence, by definition, the original observables do not exhibit single-outcome complementarity.

(iii) Full complementarity by definition is a stronger notion than single-outcome complementarity. Therefore, (ii) implies (iii). ■

APPENDIX B: FULL COMPLEMENTARITY DOES NOT IMPLY STRONG UNCERTAINTY

We will now give the example of two fine-grained projective measurements in $\mathbb{C}^5$ that do not exhibit full preparation uncertainty even though they are fully complementary. We consider two orthonormal bases (for brevity we write unnormalized vectors)

$$\begin{aligned}&|\psi_1\rangle = |0\rangle,\ |\psi_2\rangle = |1\rangle,\ |\psi_3\rangle = |2\rangle,\\ &|\psi_4\rangle = |3\rangle + |4\rangle,\ |\psi_5\rangle = |3\rangle - |4\rangle,\end{aligned} \tag{B1}$$

and

$$|\phi_1\rangle = |0\rangle + |1\rangle,\ |\phi_2\rangle = |0\rangle - |1\rangle + |2\rangle,\ |\phi_3\rangle = |3\rangle + |\chi\rangle,$$
$$|\phi_4\rangle = |4\rangle,\ |\phi_5\rangle = |3\rangle - |\chi\rangle, \tag{B2}$$

where $|\chi\rangle = (|0\rangle - |1\rangle - 2|2\rangle)/\sqrt{6}$. One readily checks that the following coarse grainings,

$$P_1 = \sum_{i=1}^{3} |\psi_1\rangle\langle\psi_1|,\ P_2 = \sum_{i=4}^{5} |\psi_1\rangle\langle\psi_1|,$$
$$Q_1 = \sum_{i=1}^{3} |\phi_1\rangle\langle\phi_1|,\ Q_2 = \sum_{i=4}^{5} |\phi_1\rangle\langle\phi_1|, \tag{B3}$$

do not exhibit uncertainty, as the input state $|\psi\rangle = (1/\sqrt{5})(2|0\rangle + |2\rangle)$ gives a deterministic outcome for both (now binary) measurements. Specifically, this state gives with certainty the outcomes corresponding to projector P_1 and Q_1, respectively. On the other hand, arbitrary coarse graining of the fine-grained measurements lead to noncommuting projectors and therefore by [43] are jointly nonmeasurable projective measurements. Hence the above two measurements, although they do not exhibit strong uncertainty, are fully complementary.

APPENDIX C: FOR NONEXTREMAL OBSERVABLES INDEPENDENCE DOES NOT IMPLY COMPLEMENTARITY

In this section, we argue that independence is not a good indicator of complementarity for nonextremal observables. Particularly, we provide an example where the independence increases under taking a convex mixture of observables. Consider a theory containing three 3-outcome observables X_1, X_2, Y whose statistics sets originate from convex combinations of three preparations P_1, P_2, P_3 such that

$$\mathbf{q}_{X_1}(P_1) = (1,0,0),\quad \mathbf{q}_{X_1}(P_2) = (0,1,0),$$
$$\mathbf{q}_{X_1}(P_3) = (0,0,1),\quad \mathbf{q}_{X_2}(P_1) = \left(\frac{1}{4},0,\frac{3}{4}\right),$$
$$\mathbf{q}_{X_2}(P_2) = \left(\frac{3}{4},0,\frac{1}{4}\right),\quad \mathbf{q}_{X_2}(P_3) = (0,1,0),$$
$$\mathbf{q}_{Y}(P_1) = \left(\frac{1}{4},\frac{3}{4},0\right), \mathbf{q}_{Y}(P_2) = \left(\frac{3}{4},\frac{1}{4},0\right),$$
$$\mathbf{q}_{Y}(P_3) = (0,0,1). \tag{C1}$$

We can verify that there exists two left-stochastic maps,

$$\Lambda_1 = \begin{bmatrix} \frac{1}{4} & \frac{3}{4} & 0 \\ \frac{3}{4} & \frac{1}{4} & 0 \\ 0 & 0 & 1 \end{bmatrix},\ \Lambda_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \tag{C2}$$

for which $\mathbf{q}_Y(P_i) = \Lambda_1 \mathbf{q}_{X_1}(P_i) = \Lambda_2 \mathbf{q}_{X_2}(P_i)$, thereby $X_{1,2} \to Y$. Consider another observable X as a convex mixture of X_1 and X_2 with equal probability. From (C1) we obtain

$$\mathbf{q}_X(P_1) = \left(\frac{5}{8},0,\frac{3}{8}\right),\quad \mathbf{q}_X(P_2) = \left(\frac{3}{8},\frac{1}{2},\frac{1}{8}\right),$$
$$\mathbf{q}_X(P_3) = \left(0,\frac{1}{2},\frac{1}{2}\right). \tag{C3}$$

Let us assume there exists a left-stochastic map,

$$\Lambda = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix},$$

such that $\Lambda \mathbf{q}_X(P_i) = \mathbf{q}_Y(P_i)$. From (C1)–(C3) we see that $\Lambda \mathbf{q}_X(P_3) = \mathbf{q}_Y(P_3)$ implies $t_{22} = t_{23} = 0$. Further, imposing this condition on $\Lambda \mathbf{q}_X(P_1) = \mathbf{q}_Y(P_1)$, we obtain $\frac{5}{8}t_{21} = \frac{3}{4}$ which implies $t_{21} = \frac{6}{5} > 1$. This is not possible for a left-stochastic map Λ. Similarly, if we assume $\Lambda \mathbf{q}_Y(P_i) = \mathbf{q}_X(P_i)$, we can check that $\Lambda \mathbf{q}_Y(P_1) = \mathbf{q}_X(P_1)$ implies $t_{21} = t_{22} = 0$; however $\Lambda \mathbf{q}_Y(P_2) = \mathbf{q}_X(P_2)$ suggests $\frac{3}{4}t_{21} + \frac{1}{4}t_{22} = \frac{1}{2}$. Hence, such a stochastic map does not exist. In other words, independence of X, Y is nonzero.

APPENDIX D: PROOF OF THE OPTIMAL SUCCESS PROBABILITY IN RANDOM ACCESS CODE FOR TWO PROJECTIVE MEASUREMENTS

We consider two quantum projective measurements corresponding to the bases $X_1 = \{|i\rangle\}_{i=1}^d$ and $X_2 = \{|\psi\rangle_j\}_{j=1}^d$ accessed by Bob. Given Alice's input a_1a_2 and her encoding state $\rho_{a_1a_2}$, the success probability of guessing a_y is

$$\sum_b p(a_b|a,b) = \mathrm{tr}\big(\big(|a_1\rangle\langle a_1| + |\psi_{a_2}\rangle\langle\psi_{a_2}|\big)\rho_{a_1a_2}\big). \tag{D1}$$

Since the operator $|a_1\rangle\langle a_1| + |\psi_{a_2}\rangle\langle\psi_{a_2}|$ is Hermitian, its eigenvectors span d-dimensional space. The optimal value of the right-hand side (D1) is the maximum eigenvalue of this operator and $\rho_{a_1a_2}$ is the corresponding eigenvector. A simple calculation leads to the fact that the maximum eigenvalue of $|a_1\rangle\langle a_1| + |\psi_{a_2}\rangle\langle\psi_{a_2}|$ is $1 + |\langle a_1|\psi_{a_2}\rangle|$. Subsequently, the average success probability (31) is

$$p_s(X_1, X_2) = \frac{1}{2} + \frac{1}{2d^2}\sum_{a_1,a_2} |\langle a_1|\psi_{a_2}\rangle|. \tag{D2}$$

To show that the above expression is the optimal success probability given the two measurements X_1, X_2, we need to show that any classical postprocessing of the outcome statistics will not yield higher success probability. Any postprocessing can be represented by the set of positive operators $\{M_{a_1}\}_{a_1=1}^d$ and $\{M_{a_2}\}_{a_2=1}^d$, corresponding to $y = 1, 2$, respectively, as follows:

$$M_{a_1} = \sum_{i=1}^{d} p(a_1|i)|i\rangle\langle i|,\ M_{a_2} = \sum_{j=1}^{d} q(a_2|j)|\psi_j\rangle\langle\psi_j|, \tag{D3}$$

for some probability distributions such that $\forall i, j,\ \sum_{a_1} p(a_1|i) = \sum_{a_2} q(a_2|j) = 1$. Since $M_{a_1} + M_{a_2}$ is a positive operator, following the previous argument we know the optimal success probability for this strategy is

$$p_s = \frac{1}{2d^2}\sum_{a_1,a_2}(||M_{a_1} + M_{a_2}||), \tag{D4}$$

where $||M||$ denotes the operator norm. Using the inequality $||X + Y|| \leqslant \max(||X||, ||Y||) + ||\sqrt{X}\sqrt{Y}||$ derived by Kittaneh [60] and the fact that $||X + Y|| \leqslant ||X|| + ||Y||$, we obtain the following relation,

$$\begin{aligned}
p_s &= \frac{1}{2d^2}\sum_{a_1,a_2}\left(\left\|M_{a_1}+M_{a_2}\right\|\right)\\
&\leqslant \frac{1}{2d^2}\sum_{a_1,a_2}\left[\max(||M_{a_1}||,||M_{a_2}||)+||\sqrt{M_{a_1}}\sqrt{M_{a_2}}||\right]\\
&\leqslant \frac{1}{2d^2}\sum_{a_1,a_2}\left[1+\left\|\sum_{i,j}\sqrt{p(a_1|i)}\sqrt{q(a_2|j)}|i\rangle\langle i|\psi_j\rangle\langle\psi_j|\right\|\right]\\
&\leqslant \frac{1}{2}+\frac{1}{2d^2}\sum_{a_1,a_2}\sum_{i,j}\sqrt{p(a_1|i)}\sqrt{q(a_2|j)}||\,|i\rangle\langle i|\psi_j\rangle\langle\psi_j|\,||\\
&\leqslant \frac{1}{2}+\frac{1}{2d^2}\sum_{a_1,a_2}\sum_{i,j}p(a_1|i)q(a_2|j)|\langle i|\psi_j\rangle|\\
&= \frac{1}{2}+\frac{1}{2d^2}\sum_{i,j}|\langle i|\psi_j\rangle|, \qquad \text{(D5)}
\end{aligned}$$

which is the same as the left-hand side of (D2). In the above derivation, we have used the fact that $\sqrt{M_{a_1}}=\sum_{i=1}^d\sqrt{p(a_1|i)}|i\rangle\langle i|$, $\sqrt{M_{a_2}}=\sum_{j=1}^d\sqrt{q(a_2|j)}|\psi_j\rangle\langle\psi_j|$.

APPENDIX E: PROOF OF EXCLUSION RELATION FROM RANDOM ACCESS CODE

Lemma 2 (quantum-mechanical uncertainty relation for exclusion-like quantity defined in terms of random access code). Consider a d-dimensional quantum system and let $X=\{|i\rangle\}_{i=1}^d$ and $Y=\{|\psi_i\rangle\}_{i=1}^d$ be two projective measurements in $\mathbb{C}^d$. Let $\mathrm{E}(S_{X,Y})$ be the exclusion-like quantum-mechanical quantity given in (43) and let $\mathrm{Ind}(S_{X,Y})$ be the quantum-mechanical complementarity measure based on the average success probability in the random access code given in (41). Then, the following uncertainty relation holds:

$$\mathrm{E}(S_{X,Y})\geqslant\frac{\mathrm{Ind}^2(S_{X,Y})}{4d}. \qquad \text{(E1)}$$

Proof. In what follows we will use the notation $U_{ij}=\langle i|\psi_j\rangle$. Recall the explicit formulas for $\mathrm{E}(S_{X,Y})$ and $\mathrm{Ind}(S_{X,Y})$,

$$\mathrm{E}(S_{X,Y})=\tfrac{1}{2}\left(1-\tfrac{1}{d}\max_\pi\textstyle\sum_{i=1}^d|U_{i\pi(i)}|\right), \qquad \text{(E2)}$$

$$\mathrm{Ind}(S_{X,Y})=\tfrac{1}{d-1}\left(\tfrac{1}{d}\textstyle\sum_{i,j=1}^d|U_{ij}|-1\right). \qquad \text{(E3)}$$

Let $p_{\max}(i):=\max_j|U_{ij}|^2$, where the maximum is over $j\in\{1,\dots,d\}$. Our proof strategy is to show that $I(S_{X,Y})>0$ implies $p_{\max}(i_0)<1$ for some i_0. As we will prove later the latter condition can be used to find the lower bound on the exclusivity $\mathrm{E}(S_{X,Y})$. By reformulating Eq. (E3) we obtain

$$\sum_{i,j=1}^d|U_{ij}|=(d-1)dI(S_{X,Y})+d, \qquad \text{(E4)}$$

from which we can readily deduce that for some i_0 we have the inequality $\sum_{j=1}^d|U_{i_0j}|\geqslant1+(d-1)\mathrm{Ind}(S_{X,Y})$. The left-hand side of this inequality can be upper bounded as

$$\sum_{j=1}^d|U_{i_0j}|\leqslant\sqrt{p_{\max}(i_0)}+\sqrt{d-1}\sqrt{1-p_{\max}(i_0)}, \qquad \text{(E5)}$$

where we have used the Shur concavity of the square-root function and the fact that for fixed i numbers $|U_{ij}|^2$ form a probability distribution. Combining (E5) with the earlier bound gives

$$\sqrt{p_{\max}(i_0)}+\sqrt{d-1}\sqrt{1-p_{\max}(i_0)}\geqslant1+(d-1)\mathrm{Ind}(S_{X,Y}). \qquad \text{(E6)}$$

Importantly, the above inequality implies that $p_{\max}(i_0)<1$ whenever $\mathrm{Ind}(S_{X,Y})>0$. To get a nontrivial upper bound on $p_{\max}(i_0)$ we apply the inequality [61] $\sqrt{p_{\max}(i_0)}\leqslant1-(1/2)[1-p_{\max}(i_0)]$, which finally gives

$$\sqrt{d-1}y-(1/2)y^2\geqslant(d-1)\mathrm{Ind}(S_{X,Y}), \qquad \text{(E7)}$$

for $y=\sqrt{1-p_{\max}(i_0)}$. By neglecting the quadratic we obtain

$$p_{\max}(i_0)\leqslant1-(d-1)\mathrm{Ind}^2(S_{X,Y}). \qquad \text{(E8)}$$

To conclude that we prove a lower bound on E in terms of $p_{\max}(i_0)$ we note that the following inequalities hold true:

$$\max_\pi\sum_i|U_{i\pi(i)}|\leqslant\sum_{i=1}^d\sqrt{p_{\max}(i)}\leqslant(d-1)+\sqrt{p_{\max}(i_0)}. \qquad \text{(E9)}$$

Plugging this bound into (E2) gives $\mathrm{E}(S_{X,Y})\geqslant1-\sqrt{p_{\max}(i_0)}/(2d)$. Together with (E8) this gives

$$\mathrm{E}(S_{X,Y})\geqslant\frac{1}{2d}\left[1-\sqrt{1-(d-1)\mathrm{Ind}\left(S^2_{X,Y}\right)}\right]. \qquad \text{(E10)}$$

Using inequality $1-\sqrt{1-x}\geqslant x/2$ valid for all $x\in(0,1)$ we obtain the final result

$$\mathrm{E}(S_{X,Y})\geqslant\frac{\mathrm{Ind}^2(S_{X,Y})}{4d}. \qquad \text{(E11)}$$

Of course, since the considered observables are clean and extremal, Ind is the same as complementarity. ■

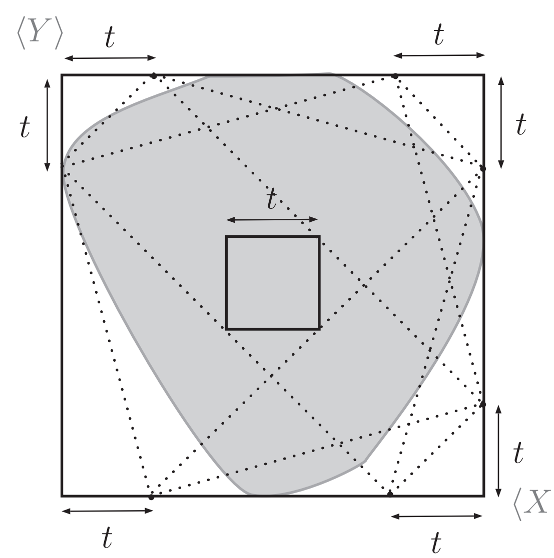

FIG. 13. We assume that the minimum distance between the corners and the points on $S_{X,Y}$, which lie on the boundary of S, is t. This, together with the fact that $S_{X,Y}$ touches all four edges of the square, imposes that the boundaries of $S_{X,Y}$ cannot be closer to the center than the dotted lines presented here. This implies that a square of length t will fit inside $S_{X,Y}$.

APPENDIX F: PROOF OF REVERSE UNCERTAINTY RELATION

For S being square for two clean and sharp observables, any theory must satisfy the reverse PUR given by $2\mathrm{C}_r \geqslant \mathrm{U}$.

Proof. Since the observables are sharp, the statistics set $S_{X,Y}$ touches all four edges of the square S. Let us say that the minimum distance between the corners and the points belonging to $S_{X,Y}$, which lie on the boundary of S, is t (see Fig. 13). It is clear from the definition that the rescaling measure of uncertainty of that point is t, and hence the uncertainty measure of the statistics set $\mathrm{U} \leqslant t$. Now, consider a square of length t taking the same origin of S. As described in Fig. 13, this square should always fit inside $S_{X,Y}$. This leads to the fact that $\mathrm{C}_r \geqslant t/2$, and subsequently $2\mathrm{C}_r \geqslant \mathrm{U}$. ■

[1] N. Bohr, Nature (London) **121**, 580 (1928).
[2] W. Heisenberg, Z. Phys. **43**, 172 (1927).
[3] P. Busch, T. Heinonen, and P. Lahti, Phys. Rep. **452**, 155 (2007).
[4] I. Białynicki-Birula and J. Mycielski, Commun. Math. Phys. **44**, 129 (1975).
[5] I. Bialynicki-Birula and L. U. Rudnicki, in *Statistical Complexity: Applications in Electronic Structure*, edited by K. Sen (Springer Netherlands, Dordrecht, 2011), pp. 1–34.
[6] M. H. Partovi, Phys. Rev. A **84**, 052117 (2011).
[7] S. Friedland, V. Gheorghiu, and G. Gour, Phys. Rev. Lett. **111**, 230401 (2013).
[8] Z. Puchala, L. Rudnicki, and K. Zyczkowski, J. Phys. A: Math. Theor. **46**, 272002 (2013).
[9] V. Giovannetti, Phys. Rev. A **70**, 012102 (2004).
[10] O. Gühne and M. Lewenstein, Phys. Rev. A **70**, 022316 (2004).
[11] S. P. Walborn, B. G. Taketani, A. Salles, F. Toscano, and R. L. de Matos Filho, Phys. Rev. Lett. **103**, 160505 (2009).
[12] P. J. Coles, M. Berta, M. Tomamichel, and S. Wehner, Rev. Mod. Phys. **89**, 015002 (2017).
[13] G. Ver Steeg and S. Wehner, Quantum Inf. Comput. **9**, 0801 (2009).
[14] J. Oppenheim and S. Wehner, Science **330**, 1072 (2010).
[15] R. Ramanathan, D. Goyeneche, S. Muhammad, P. Mironowicz, M. Grünfeld, M. Bourennane, and P. Horodecki, Nat. Commun. **9**, 4244 (2018).
[16] E. Hänggi and S. Wehner, Nat. Commun. **4**, 1670 (2013).
[17] M. Plávala, Phys. Rev. A **94**, 042108 (2016).
[18] A. Jenčová and M. Plávala, Phys. Rev. A **96**, 022113 (2017).
[19] P. Janotta and H. Hinrichsen, J. Phys. A: Math. Gen. **47**, 323001 (2014).
[20] P. Busch, T. Heinosaari, J. Schultz, and N. Stevens, Europhys. Lett. **103**, 10002 (2013).
[21] S. N. Filippov, T. Heinosaari, and L. Leppäjärvi, Phys. Rev. A **95**, 032127 (2017).
[22] N. Bohr, *Philosophical Writings of Niels Bohr*, Vol. 2 (Ox Bow Press, Woodbridge, CT, 1949).
[23] See also [37] for a comprehensive account of the evolution of Bohr's views on the notions of uncertainty and complementarity.
[24] R. W. Spekkens, Phys. Rev. A **71**, 052108 (2005).
[25] M. S. Leifer, Quanta **3**, 67 (2014).
[26] V. Narasimhachar, A. Poostindouz, and G. Gour, New J. Phys. **18**, 033019 (2016).
[27] F. Buscemi, M. Keyl, G. M. D'Ariano, P. Perinotti, and R. F. Werner, J. Math. Phys. **46**, 082109 (2005).
[28] E. Haapasalo, T. Heinosaari, and J.-P. Pellonpaa, Quant. Inf. Process. **11**, 1751 (2012).
[29] M. Oszmaniec, L. Guerini, P. Wittek, and A. Acín, Phys. Rev. Lett. **119**, 190501 (2017).
[30] M. Oszmaniec, F. B. Maciejewski, and P. Zbigniew, Phys. Rev. A **100**, 012351 (2019).
[31] L. Guerini, J. Bavaresco, M. Terra Cunha, and A. Acín, J. Math. Phys. **58**, 092102 (2017).
[32] S. N. Filippov, T. Heinosaari, and L. Leppäjärvi, Phys. Rev. A **97**, 062102 (2018).
[33] E. H. Kennard, Z. Phys. **44**, 326 (1927).
[34] H. P. Robertson, Phys. Rev. **34**, 163 (1929).
[35] D. Deutsch, Phys. Rev. Lett. **50**, 631 (1983).
[36] H. Maassen and J. B. M. Uffink, Phys. Rev. Lett. **60**, 1103 (1988).
[37] A. Plotnitsky, Phys. Scr. **2014**, 014002 (2014).
[38] D. Petz, Rep. Math. Phys. **59**, 209 (2007).
[39] P. Busch, P. Lahti, and R. F. Werner, J. Math. Phys. **55**, 042111 (2014).
[40] M. J. W. Hall, Phys. Rev. Lett. **74**, 3307 (1995).
[41] A. Grudka, M. Horodecki, P. Horodecki, R. Horodecki, W. Klobus, and L. Pankowski, Phys. Rev. A **88**, 032106 (2013).
[42] P. J. Coles and M. Piani, Phys. Rev. A **89**, 022112 (2014).
[43] T. Heinosaari, D. Reitzner, and P. Stano, Found. Phys. **38**, 1133 (2008).
[44] P. Lahti and S. Pulmannová, Rep. Math. Phys. **39**, 339 (1997).
[45] E. Haapasalo, J.-P. Pellonpää, and R. Uola, Lett. Math. Phys. **105**, 661 (2015).
[46] M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Żukowski, Nature (London) **461**, 1101 (2009).
[47] L. Czekaj, M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. A **95**, 022119 (2017).
[48] G. Brassard, H. Buhrman, N. Linden, A. A. Méthot, A. Tapp, and F. Unger, Phys. Rev. Lett. **96**, 250401 (2006).
[49] M. Navascues and H. Wunderlich, Proc. R. Soc. London A **466**, 881 (2010).
[50] J. Lodyga, W. Klobus, R. Ramanathan, A. Grudka, M. Horodecki, and R. Horodecki, Phys. Rev. A **96**, 012124 (2017).
[51] A. Ambainis, D. Leung, L. Mančinska, and M. Ozols, arXiv:0810.2937.
[52] M. Czechlewski, D. Saha, A. Tavakoli, and M. Pawłowski, Phys. Rev. A **98**, 062305 (2018).
[53] E. A. Aguilar, J. J. Borkała, P. Mironowicz, and M. Pawłowski, Phys. Rev. Lett. **121**, 050501 (2018).
[54] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).
[55] We remark that steering and joint nonmeasurability are shown to be equivalent in quantum theory [58, 59].

[56] E. Bagan, J. Calsamiglia, J. A. Bergou, and M. Hillery, Phys. Rev. Lett. **120**, 050402 (2018).
[57] L.-L. Sun, S. Yu, and Z.-B. Chen, arXiv:1808.06416.
[58] M. T. Quintino, T. Vértesi, and N. Brunner, Phys. Rev. Lett. **113**, 160402 (2014).
[59] R. Uola, T. Moroder, and O. Gühne, Phys. Rev. Lett. **113**, 160403 (2014).
[60] F. Kittaneh, J. Funct. Analysis **143**, 337 (1997).
[61] This follows from the simple inequality $\sqrt{1-y} \leqslant 1-(y/2)$ used for $y = 1 - p_{\max}(i_0)$.