Information Invariance and Quantum Probabilities

Časlav Brukner · Anton Zeilinger

Abstract We consider probabilistic theories in which the most elementary system, a two-dimensional system, contains one bit of information. The bit is assumed to be contained in any complete set of mutually complementary measurements. The requirement of *invariance* of the information under a *continuous* change of the set of mutually complementary measurements uniquely singles out a measure of information, which is *quadratic* in probabilities. The assumption which gives the same scaling of the number of degrees of freedom with the dimension as in quantum theory follows essentially from the assumption that all physical states of a higher dimensional system are those and only those from which one can post-select physical states of two-dimensional systems. The requirement that no more than one bit of information (as quantified by the quadratic measure) is contained in all possible post-selected two-dimensional systems is equivalent to the positivity of density operator in quantum theory.

1 Introduction: Failure of Classical Concepts

In general quantum mechanics only makes probabilistic predictions for individual events. Can one go beyond quantum mechanics in this respect? With an aim to ar-

---

Č. Brukner (✉) · A. Zeilinger
Institute for Quantum Optics and Quantum Information, Austrian Academy of Sciences,
Boltzmanngasse 3, 1090, Vienna, Austria
e-mail: caslav.brukner@univie.ac.at

Č. Brukner · A. Zeilinger
Faculty of Physics, University of Vienna, Boltzmanngasse 5, 1090, Vienna, Austria

gue for incompleteness of the quantum-mechanical description Einstein, Podolsky and Rosen [1] (EPR) introduced the notions of "locality" and "elements of physical reality" in their seminal paper from 1935. In the EPR words the two notions read: (Locality) "Since at the time of measurements the two systems no longer interact, no real change can take place in the second system in consequence of anything that may be done to the first system"; and (Elements of physical reality) "If, without in any way disturbing a system, we can predict with certainty (i.e., with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity."

The theorem of Greenberger, Horne and Zeilinger [2, 3] (GHZ) showed that the mere concept of existence of local "elements of physical reality" as introduced by EPR is in a contradiction with quantum mechanical predictions. Which—if not both—of the two EPR premisses is violated is an open question. One could, for example, relax the premiss of "elements of physical reality" and consider local probabilistic (also called stochastic) theories in which the individual local result is not assumed to be pre-determined, but only its probability to occur. To discuss what it explicitly means consider two space-like separated parties, colloquially called Alice and Bob, who perform measurements in their local laboratories. In every experimental run one defines the conditional probability (describing correlations) $p(a, b|x, y, \lambda)$ that Alice's and Bob's outcomes are a and b, given that their (free [4]) choices of measurement settings are x and y, respectively. Here λ is the full set of, hidden and not hidden, variables that may include all the information about the past of both Alice and Bob except for their choices of measurement settings. In a local probabilistic hidden-variable theory one requires that $p(a, b|x, y, \lambda) = p(a|x, \lambda) \cdot p(b|y, \lambda)$ is satisfied, such that measurable conditional probability becomes $p(a, b|x, y) = \int d\lambda \rho(\lambda) p(a|x, \lambda) \cdot p(b|y, \lambda)$, where $\rho(\lambda)$ is the probability distribution of the hidden variable. The theorem of Bell [5], which historically precedes the GHZ theorem, shows that not even a local probabilistic hidden-variable theory can agree with all predictions of quantum theory. What is the lesson we can learn from this? As we here speak about theories that treat probabilities as irreducible and which are local, but nonetheless contradict quantum predictions, for some authors this indicates that nature is non-local.

While the mere existence of Bohm's model [6] demonstrates that non-local hidden-variables are a logically valid option, we now know that there are plausible models, such as Leggett's crypto-nonlocal hidden-variable model [7], that are in disagreement with both quantum predictions and experiment [8, 9]. But, perhaps more importantly in our view is that if one is ready to consider probabilistic theories, then there is no immediate reason to require the locality condition in the form given above. Violation of this condition is not in conflict with the theory of relativity, as it does not imply the possibility of signalling superluminally. To the contrary, quantum correlations cannot be used to communicate from Alice to Bob, nor from Bob to Alice, but do violate Bell's inequalities.[1] It is therefore legitimate to consider quantum theory as

---

[1]In relation to this it is symptomatic that quantum formalism makes no difference between the description of two (space-like) separated particles and of two degrees of freedom of a single particle with a corresponding Hilbert space dimension. This suggests that spatial distance in the ordinary three-dimensional

a probability theory subject to or even derivable from more general principles such as the principle of no-signalling [10, 11] or an information-theoretical principle [12–14].

Whether or not one may be able to resolve the question which (if not both) of the two EPR premises is violated in quantum mechanics, one can ask how plausible the hidden-variable program is on its own. A particularly interesting question is what is the amount of resources in terms of the number of hidden-variable states a hidden-variable model must consume to be in agreement with quantum theory. In papers of Hardy [15], Montina [16] and Dakic *et al.* [17] it was shown that in the limit of large number of measurements no hidden-variable theory that agrees with quantum predictions could use less hidden-variable states than the straightforward "brute force" model in which every quantum state is associated with one such hidden state. This implies that no hidden-variable theory can provide a description that is more efficient than quantum theory itself. Even for the simplest quantum system like electron spin or photon polarization, any hidden-variable approach is extremely resource demanding, requiring infinitely many hidden-variable states in order to fulfill the requirement of explaining all possible measurements. It is interesting here to contrast this finding with Feynman's words: "It always bothers me that, according to the laws as we understand them today, it takes a computing machine an infinite number of logical operations to figure out what goes on in no matter how tiny a region of space and no matter how tiny a region of time, ... why should it take an infinite amount of logic to figure out what one tiny piece of space-time is going to do?"

Concluding the introduction we note that while maintaining the hidden-variable program and giving up locality may logically be possible, the analysis given above in our opinion supports Bohr's view of [18] "the necessity of a final renunciation of the classical ideal of causality and a radical revision of our attitude towards the problem of physical reality."

2 A Foundational Program Based on Probabilities

Recently, several authors have been looking at ways of deriving quantum theory from reasonable principles that are motivated by putting primacy on the concept of information or on the concept of probability which again can be seen as a way of quantifying information. Particularly interesting papers which discuss this idea are Refs. [11–14, 19–36]. This incomplete list of various approaches should suggest that it is possible to derive a large part of the quantum formalism without recourse to *ad hoc* assumptions about the abstract Hilbert space description of quantum systems. Rather, a goal is that the usual rather formal mathematical axiomatization (see, for example, Ref. [37]) arises as a consequence.

In this paper we consider probabilistic theories built upon the physical principle of limited information content of any system. The paper certainly does not present a complete axiomatic derivation of quantum theory but rather it demonstrates that some essential elements of the theory can be obtained from the concept of information.

---

space is irrelevant in the abstract Hilbert space description of quantum mechanical systems. In our view this relativizes the violation of locality condition as a possible explanation of Bell's theorem.

We also discuss how our approach relates to others, like those of Hardy [25] and Aaronson [38].

In the probabilistic theory the state of a system is defined by the set of probabilities associated with every measurement that may be performed on the system. Given a list of the probabilities associated with every conceivable measurement, labeled by i, one could write a state of a system as a probability vector: $\mathbf{p} = (\ldots, p_i, \ldots)$. The list might contain superfluous information, and instead one can look for a list that contains the minimal number of probabilities that are sufficient to fully specify the state in the sense that the probability of any possible result for any conceivable measurement can be derived therefrom.

Hardy [25] discussed two important integers in the reconstruction of such theories. The first one is the smallest number K of different measurements that specify the state completely. The second one is the dimension N which is the maximum number of states that can be distinguished from one another in a single shot experiment. Two states are distinguishable if there exists a choice of measurement setting for which the sets of possible outcomes they can give rise to are disjoint. Therefore, the state is represented by a probability vector $\mathbf{p} = (p_1, \ldots, p_K)$. Wootters [39] and Hardy [25] gave the parameter counting argument for composite systems to determine the function $K(N)$. The argument is based on the following assumption:

(i) (*Composite Systems*) Upon combining constituents into a composite system, dimension N and number K of degrees of freedom are multiplicative, i.e. $K(N_1 N_2) = K(N_1)K(N_2)$, where N_i, $i = 1, 2$, is dimension and $K(N_i)$ the degree of freedom of the i-th constituent.

The condition leads to $K(N) = N^r - 1$ with $r \in \mathbf{N}$. This suggests the existence of a hierarchy of theories, which correspond to different r [25, 40, 41].

We begin our considerations from the guiding principle that [14]

(ii) (*Limited Information*) The most elementary system (two-dimensional system) contains one bit of information.

The most elementary system is of dimension two. The principle states that there is a measurement with two distinguishable outcomes in either of which the system may give a deterministic answer. When the system gives such an answer it is called to be in a pure state. It is clear that depending on the particular function $K(N)$ one can have different theories in agreement with the principle. For $K = 2^1 - 1 = 1$ one has a classical bit for which only two states "0" and "1" satisfy the principle when the system definitely gives either answer. Any convex combination (classical mixture) of the two states, p"0"$+(1 - p)$"1", gives rise to a (mixed) state that has *less* than one bit of information. This statement is independent of how we quantify information, i.e. of the particular choice of information measure $H(p, 1 - p)$. Whenever one has a convex combination of two states, one loses the information from which state a particular sample comes from.

The next most simple case is for $K = 2^2 - 1 = 3$ which is the case in quantum theory. One needs three independent measurements to completely specify the state of a two-dimensional system. We make here a non-trivial assumption that one can choose mutually complementary measurements for these three independent measurements in a general probabilistic theory with $K = 3$ (in quantum theory these

measurements are known as mutually unbiased [42, 43]). They have the property of mutual exclusiveness: the total knowledge of one observable is at the cost of total ignorance about the other two complementary ones. More precisely, if probability for an outcome in one of the three experiments is unity, then the two outcomes in either of the two complementary experiments are equally probable. In this sense mutually complementary measurements constitute "maximally independent pieces of information" that are contained in the system. They serve as a useful "reference frame" to describe how information that is contained in a system is distributed over independent observables. An explicit example of a set of mutually complementary observables in quantum theory are three spin projections along orthogonal directions of a spin-1/2 particle. Having this example in mind we will denote three mutually complementary measurements as x, y and z. The state is thus given by $\mathbf{p} = (p_x, 1 - p_x, p_y, 1 - p_y, p_z, 1 - p_z)$. (One can additionally reduce the number of components in the vector since the probabilities sum up to unity in every individual experiment. We leave the extended form for the purposes of further discussion.)

Concluding this section we note that in a general probabilistic theory of a two-dimensional state there are $K = 2^r - 1$, $r = 1, 2, 3, \ldots$, mutually complementary measurements [41]. This number agrees with the parameter counting arguments [25, 39] mentioned above and, independently, follows from *operational* considerations in which complementary measurements reveal mutually complementary properties of black box configurations in the framework of black-box computation [41].

3 Uniqueness of the Quadratic Measure of Information

The necessary choice between observing path information and the observability of interference patterns is one of the most basic manifestations of quantum complementarity, as introduced by Niels Bohr. It does not only compromise the extreme cases of maximal knowledge of path information at the expense of complete loss of interference and vice versa, but also intermediate situations in which one can obtain some partial knowledge about the particle's path and still observe an interference pattern of reduced contrast as compared to the ideal interference situation [44–48]. This and other manifestations of quantum complementarity suggests that our total knowledge, or information, about the outcomes in complementary experiments that can be performed on a given quantum system is limited. This information may then manifest itself fully as path information or as modulation of the interference pattern or partially in both to the extent defined by the finiteness of information. The question is how to give a formal mathematical description to this observation.

Bohr [49] remarked that ". . . phenomena under different experimental conditions, must be termed complementary in the sense that each is well defined and that together they exhaust all definable knowledge about the object concerned." This suggests that the total information content of a quantum system is somehow contained in the full set of mutually complementary experiments. We follow here a natural choice and define the total information (of 1 bit) of a two-dimensional system as the *sum* of the individual measures of information over three mutually complementary experiments. For convenience we will use here not a measure of information or knowledge, but

rather its opposite, a measure of uncertainty or entropy. This choice has no relevance for our findings. Therefore, the total uncertainty about a quantum system is defined as

$$H_{total} = H(p_x, 1 - p_x) + H(p_y, 1 - p_y) + H(p_z, 1 - p_z) = 2, \qquad (1)$$

where $H(p_u, 1 - p_u)$ is some measure of uncertainty that is associated to the probability distribution observed in experiment u. At this stage of the argument it is assumed that H can be an arbitrary function of probabilities that fulfills (1) and is normalized such that for an individual experiment it takes its maximal value of 1 for completely random outcomes ($p_u = 1/2$) and minimal value of 0 for deterministic results ($p_u = 0$ or 1). Note, however, that one can have minimal uncertainty in at most one of the three mutually complementary measurements, therefore the sum in (1) has to be 2. It should also be mentioned that in general $H_{total} \geq 2$ to include the case of mixed states, which are defined as convex mixtures of pure states.

In a general probabilistic theory one can analogously define the total uncertainty as

$$H_{total} = \sum_{j=1}^{2^r - 1} H(p_j, 1 - p_j) = 2^r - 2, \qquad (2)$$

where we sum over all $2^r - 1$ mutually complementary measurements and again in at most one of them we can have full knowledge. The quantum case (1) is recovered for $r = 2$.

Which measure of uncertainty should we use to quantify uncertainties in (1) and (2)? In literature one can find a large number of different measures. In fact, there is an (uncountable) infinite number of generalized measures, known as entropies of degree α,

$$H_\alpha(p_1, \ldots, p_n) := k \frac{1 - \sum_{i=1}^n p_i^\alpha}{\alpha - 1}, \qquad (3)$$

where k and α are two constants and n is the number of different outcomes. They were introduced by Havrda-Charvát [50] and are also known as Tsallis [51] entropies in statistical physics. They are one-parameter generalization of the Shannon entropy,[2]

$$\lim_{\alpha \to 1} H_\alpha = H_1 = -k \sum_{i=1}^n p_i \log p_i. \qquad (4)$$

We will now introduce two physically reasonable requirements on α-entropies, which will finally uniquely single out the $\alpha = 2$ entropy. We require (a) the total uncertainty of a system to be independent of the particular choice of the complete set of mutually complementary measurements, and (b) that the change from one to another set can be performed in a continuous fashion. Equivalent is the requirement that

---

[2]The characteristic property of Tsallis entropies is called pseudoadditivity: $\frac{H_\alpha(AB)}{k} = \frac{H_\alpha(A)}{k} + \frac{H_\alpha(B)}{k} + (1 - \alpha) \frac{H_\alpha(A)}{k} \frac{H_\alpha(B)}{k}$ holds true for two mutually independent finite event systems A and B.

(iii) (*Information Invariance & Continuity*) The total uncertainty (or total informa-
     tion of one bit) is invariant under a continuous change between different com-
     plete sets of mutually complementary measurements.

As noted above a selection of a specific set is like a selection of a reference frame:
the total uncertainty should be independent of the observers' choice of how he rep-
resents his knowledge about the system. Mathematically, the property of invariance
means that (1) is valid for all sets of choices of x, y and z. From the form of α-
entropies (3) one concludes that the invariance property is equivalent to the condi-
tion that the α-*norm of the probability vector* **p** *is preserved under the change of
the set of mutually complementary measurements*. The α norm of a given vector
$\mathbf{r} = (r_1, \ldots, r_m)$ is defined as $\|\mathbf{r}\|_\alpha := (\sum_{i=1}^m |r_i|^\alpha|)^{1/\alpha}$.

By choosing different set of complementary measurements, the state **p** will be
transformed. Hence, it will go from **p** to some new state **f(p)**, where **f** is a vector
function associated with the transformation. Hardy [25] showed that the **f** is actually
a linear transformation. Hence the transformation is given by $\mathbf{p} \to \hat{A}\mathbf{p}$, where A is a
real matrix.[3]

We next use a result of Aaronson [38], which is derived in a different context.
He considered which matrices $\hat{A} \in \mathbf{R}^{m \times m}$ have the property that for *all* vectors **r**,
$\|\hat{A}\mathbf{r}\|_\alpha = \|\mathbf{r}\|_\alpha$. He proved that *if $\alpha \neq 2$, then the only α-norm-preserving linear
transformations are permutations of diagonal matrices*. These transformations are
discrete and cannot account for a *continuous* change of vector **r** that would preserve
its norm. If $\alpha = 2$, the norm-preserving transformations are orthogonal matrices that
continuously change **r**.

Taking now the probability vector **p** in the proof of Ref. [38] one can show that
the only measure of information that satisfies the requirement of invariance under
continuous change of the probability vector is the entropy of degree $\alpha = 2$. One
should here be careful, because the probability vectors are not arbitrarily 6-vectors
as in the proof of Ref. [38] but of the form $(p_x, 1 - p_x, p_y, 1 - p_y, p_z, 1 - p_z)$ (the
probabilities are nonnegative and they sum up to unit in any of the three individual
experiments). This implies that the transformation must be *stochastic* in the sectors
$(p_x, 1 - p_x)$, $(p_y, 1 - p_y)$ and $(p_z, 1 - p_z)$ of the probability vector, which corre-
spond to individual experiments x, y and z, respectively. With this additional con-
dition fulfilled one can show that the original statement is still valid: if $\alpha \neq 2$, then
the only α-norm-preserving linear transformations of probability vectors are permu-
tations of diagonal matrices. To guarantee that the transformation is stochastic in the
corresponding sectors of the probability vector one has that two probabilities from
one sector $(p_i, 1 - p_i)$ are always permuted with two probabilities from the second
sector $(p_j, 1 - p_j)$. An explicit example of one such discrete transformation is as

---

[3]It should be noted that non-linear transformations could lead to superluminal signaling [52], violation of
the second law of thermodynamics [53] or solubility of NP-complete problems [54], which all question
their physical justification.

follows:

$$\hat{A} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{5}$$

It performs the following mapping: $p_x \to p_y$, $p_y \to 1 - p_x$ and $p_z \to p_z$. We conclude that 2-entropy is the only α-entropy that is conserved under continuous transformations. The transformations are orthogonal matrices in the three-dimensional space of spin mean values $(2p_x - 1, 2p_y - 1, 2p_z - 1)$. We note that the requirement of conservation of the measure of information under continuous transformation is closely related to the continuity axiom in the formulation of Hardy [55]: "There exists a continuous reversible transformation on a system between any two pure states of that system for systems of any dimension N."

In Ref. [26, 27] we showed that 2-entropy has the property of invariance and in Ref. [56] that Shannon's measure does not have this property, but the general proof of uniqueness was lacking.

4 Higher-Dimensional Systems: Any Post-Selected Two-Dimensional System Contains at Most One Bit of Information

We now give an interesting observation on the scaling of the number of degrees of freedom $K(N)$ with dimension N. We assume that it is possible to perform a *selective* measurement on a higher-dimensional system to prepare a genuine two-dimensional system. Suppose that two-dimensional system is the most fundamental one in the following particular sense:

(iv) (*Higher-Dimensional Systems*): Allowed physical states of a general N dimensional system are those and only those from which one can generate allowed physical states of two dimensional systems in post-selected measurements (i.e. systems that contain at most one bit of information).

As mentioned above the parameter counting argument based on considerations of composite systems requires that $K(N) = N^r - 1$ with $r \in \mathbf{N}$. Suppose that a single two-dimensional state requires $K(2) := m$ independent measurements (e.g. mutually complementary ones) to be described completely. One can now think about all possible independent measurements that can be performed on all possible two-dimensional systems that can be post-selected from the full system of dimension N. From the selective measurement one obtains $N - 1$ independent probabilities (the sum of all of them is unity). For each pair (i, j) of outcomes in the selective measurement one constructs $m - 1$ additional measurements, to describe the two-dimensional system completely. This gives altogether

$$K(N) = N - 1 + \frac{1}{2}N(N - 1)(m - 1), \tag{6}$$

which is equal to $N^r - 1$ only for $r = 2$ and $m = 3$. This is the case of quantum theory.

Taking scaling $K(N) = N^2 - 1$ as given we represent the state by a probability vector whose first N components are chosen to be probabilities $p_1, \ldots, p_N$ of N distinguishable outcomes in a single measurement. This measurement will be called Z measurement. In addition, for each pair (i, j) of outcomes of the Z measurement one constructs two additional measurements X_{ij} and Y_{ij}. These two measurements have all outcomes the same as the Z measurement except for the pair of outcomes (i, j). From the two additional measurements one obtains two additional independent probabilities p_{xij} and p_{yij} for each pair (i, j). This gives altogether $N^2 - 1$ independent parameters as required. The probability vector thus can be given as $\mathbf{p} = (p_1, \ldots, p_N, \ldots, p_{xij}, p_{yij}, \ldots)$. Next, consider setting up the Z measurement with N distinguishable outcomes and placing $N - 2$ detectors to register all but the i and j outcomes as shown in the Fig. 1. One then selects all those cases when no detector registers an event. We assume that in this post-selected manner one can prepare an arbitrary two-dimensional system (elementary system). The two additional measurements described above together with the measurement with outcomes i and j constitute a complete set of mutually complementary measurements for the two-dimensional system. In the subensemble of cases when the postselection occurs the total information contained in a complete set of mutually complementary measurements in the subselected elementary system is at most one bit of information. We
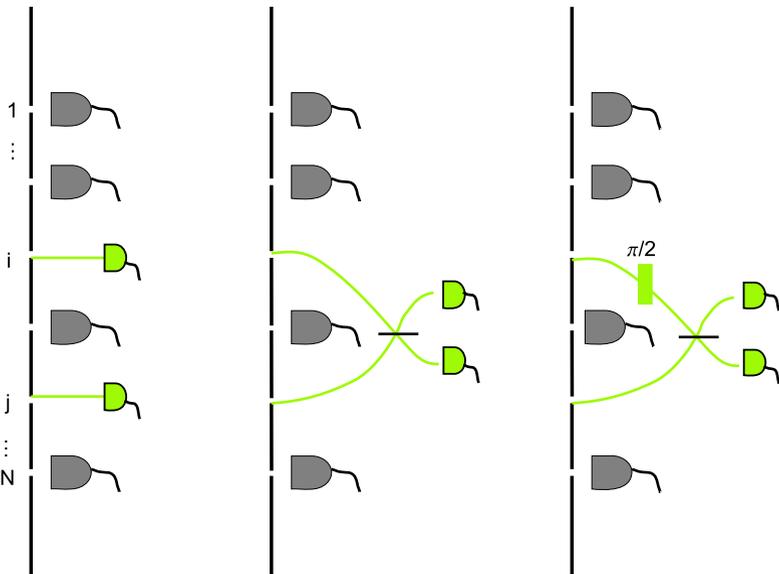

Fig. 1 Measurements to post-select a two-dimensional system. If none of the grey detectors registers an event, a two-dimensional system is prepared by post-selection from a higher-dimensional one. After post-selection one can perform a complete set of three mutually complementary measurements. In the figure the three measurements are (1) which-path measurement; (2) interference experiment with zero relative phase between two path states and (3) interference experiment with $\pi/2$ relative phase between two path states

note that this requirement is closely related to Hardy's subspace axiom [55]: "... all systems of dimension N; or systems of higher dimension but where the state is constrained to an N-dimensional subspace, have the same properties."

Since we have performed a "degenerative measurement" in which we have not distinguished between outcomes i and j it is natural to assume that the state in the two-dimensional sector does not change in the measurement except for normalization. Thus, one has $\mathbf{p} = \frac{1}{p_i+p_j}(p_i, p_j, p_{xij}, 1 - p_{xij}, p_{yij}, 1 - p_{yij})$.

It follows from (iv) that the principle of finiteness of information (ii)—the total information content of a two-dimensional system is not larger than one bit of information—should also apply to all possible two-dimensional systems that can be post-selected from a higher-dimensional system. We will show that this requirement is equivalent to the positivity of the density operator in the Hilbert space formulation of quantum theory. To prove this it is crucial to use the quadratic measure of information in the expression for the total uncertainty:

$$
\begin{aligned}
H_{total} &= H_2\left(\frac{p_i}{p_i + p_j}, \frac{p_j}{p_i + p_j}\right) + H_2\left(\frac{p_{xij}}{p_i + p_j}, 1 - \frac{p_{xij}}{p_i + p_j}\right) \\
&\quad + H_2\left(\frac{p_{yij}}{p_i + p_j}, 1 - \frac{p_{yij}}{p_i + p_j}\right) \\
&= \frac{4}{(p_i + p_j)^2}[p_i p_j + p_{xij}(1 - p_{xij}) + p_{yij}(1 - p_{yij})] \geq 2.
\end{aligned}
\tag{7}
$$

This is valid for any choice of measurements Z, X_{ij} and Y_{ij}. Here we put $k = 2$ in the definition of 2-entropy to guarantee the proper normalization as in (1).

We now show that a set of conditions (7) applied to all (i, j), $i \neq j \in \{1, \ldots, N\}$, and for all possible complete Z measurements is equivalent to positivity of density operators in quantum theory. We follow the idea of the proof of Ref. [57]. Consider a general Hermitian operator $\hat{\rho}$ of arbitrary dimension N with unit trace[4] $\mathrm{Tr}\{\hat{\rho}\} = 1$. The operator has real diagonal elements and due to unit trace condition there always exists at least one positive diagonal element. Suppose that we have performed the post-selected measurement as introduced above. Then the density operator of the two-dimensional system is $\hat{\rho} = \frac{1}{\rho_{ii}+\rho_{jj}}\left(\begin{smallmatrix} \rho_{ii} & \rho_{ij} \\ \rho_{ji} & \rho_{jj} \end{smallmatrix}\right)$. We choose for the set of mutually complementary measurement the three pseudo-spins: $\hat{\sigma}_z = |i\rangle\langle i| - |j\rangle\langle j|$, $\hat{\sigma}_x = |i\rangle\langle j| + |j\rangle\langle i|$ and $\hat{\sigma}_y = i(-|i\rangle\langle j| + |j\rangle\langle i|)$. Calculating the corresponding probabilities in (7) one obtains $\rho_{ii}\rho_{jj} \geq |\rho_{12}|^2 \geq 0$ for *all* choices of i and j. This with the condition of unit trace implies that *all* diagonal elements are non-negative, i.e. $\rho_{ii} \geq 0$ for all i. Since the condition on the total uncertainty applies to all possible choices of the Z measurement, this shows that the density operator is positive.

Concluding this section we note that information invariance (iii) together with the homogeneity of the parametric space leads to the well known cosine-law (Malus' law) for quantum probabilities [26, 27]. Again it is crucial to use quadratic measure of information.

---

[4]The unit trace condition can be relaxed if instead one requires a positivity of variance. See Ref. [57].

5 Conclusions

We think that there are sufficient evidences which indicate that so far hidden-variable approach could not encourage any new phenomenology that might result in the hope for a progressive research program towards answering Wheeler's famous question "Why the quantum?". We suggest that a clear and promising alternative to hidden-variable approach is to consider theories in which probabilities have irreducible character. A guiding idea that we follow here in order to impose a non-trivial structure on a probabilistic theory is the principle that *a two-dimensional system contains one bit of information.*

The assumption of invariance, i.e. that this one bit of information is conserved under the continuous change of a set mutually complementary measurements, uniquely singles out that specific measure of information which is quadratic in probabilities. This says, however, nothing about the scaling $K(N)$ of the number of degrees of freedom with the number of dimension N of a system. If, however, one requires that *all* physical states of an N dimensional system are those and only those from which one can post-select physical two-dimensional states, one obtains the scaling $K(N) = N^2 - 1$ as in quantum mechanics, taking in addition the parameter-counting argument of Wootters [39] and of Hardy [25] into account.

The principle of limited information is assumed to apply to every two-dimensional systems and therefore also to all such post-selected two-dimensional systems. Using the measure of information that is quadratic in probabilities one can show that this requirement is equivalent to the positivity of the density operator in quantum mechanics.

References

1. Einstein, A., Podolsky, B., Rosen, N.: Can quantum-mechanical description of physical reality be considered complete? Phys. Rev. **47**, 777–780 (1935)
2. Greenberger, D.M., Horne, M., Zeilinger, A.: Bell's theorem. In: Kafatos, M. (ed.) Quantum Theory, and Conceptions of the Universe, pp. 73–76. Kluwer Academic, Dordrecht (1989)
3. Greenberger, D.M., Horne, M., Shimony, A., Zeilinger, A.: Bell's theorem without inequalities Am. J. Phys. **58**, 1131–1143 (1990)
4. Bell, J.S.: Free variables and local causality. Dialectica **39**, 103–106 (1985)
5. Bell, J.S.: On the Einstein-Podolsky-Rosen paradox. Physics **1**, 195–200 (1964). Reprinted in J.S. Bell, Speakable and Unspeakable in Quantum Mechanics, Cambridge Univ. Press, Cambridge (1987)
6. Bohm, D.: A suggested interpretation of the quantum theory in terms of "hidden" variables, I and II. Phys. Rev. **85**, 166–193 (1952)
7. Leggett, A.J.: Nonlocal hidden-variable theories and quantum mechanics: an incompatibility theorem. Found. Phys. **33**, 1469 (2003)
8. Gröblacher, S., Paterek, T., Kaltenbaek, R., Brukner, Č, Zukowski, M., Aspelmeyer, M., Zeilinger, A.: An experimental test of non-local realism. Nature **446**, 871–875 (2007)
9. Branciard, C., Ling, A., Gisin, N., Kurtsiefer, C., Lamas-Linares, A., Scarani, V.: Experimental falsification of Leggett's nonlocal variable model. Phys. Rev. Lett. **99**, 210407 (2007)
10. Popescu, S., Rohrlich, D.: Quantum nonlocality as an axiom. Found. Phys. **24**, 379 (1994)

11. Barrett, J.: Information processing in generalized probabilistic theories. Phys. Rev. A **75**, 032304 (2007)
12. von Weizsäcker, C.F.: Aufbau der Physik. Hanser, Munich (1985). In German
13. Wheeler, J.A.: Information, physics, quantum: The search for links. In: Zurek, W. (ed.) Complexity, Entropy, and the Physics of Information. Addison-Wesley, Reading (1990)
14. Zeilinger, A.: A foundational principle for quantum mechanics. Found. Phys. **29**, 631–643 (1999)
15. Hardy, L.: Quantum ontological excess baggage. Stud. Hist. Philos. Mod. Phys. **35**, 267 (2004)
16. Montina, A.: Exponential growth of the ontological space dimension with the physical size. Phys. Rev. A **77**, 022104 (2008)
17. Dakic, B., Suvakov, M., Paterek, T., Brukner, Č.: Efficient hidden-variable simulation of measurements in quantum experiments. Phys. Rev. Lett. **101**, 190402 (2008)
18. Bohr, N.: Can quantum-mechanical description of physical reality be considered complete? Phys. Rev. **48**, 696 (1935)
19. Wootters, W.K.: Statistical distance and Hilbert space. Phys. Rev. D **23**, 357 (1981)
20. Fivel, D.I.: How interference effects in mixtures determine the rules of quantum mechanics. Phys. Rev. A **59**, 2108 (1994)
21. Summhammer, J.: Maximum predictive power and the superposition principle. Int. J. Theor. Phys. **33**, 171 (1994)
22. Summhammer, J.: Quantum theory as efficient representation of probabilistic information. arXiv:quant-ph/0701181
23. Bohr, A., Ulfbeck, O.: Primary manifestation of symmetry. Origin of quantal indeterminacy. Rev. Mod. Phys. **67**, 1–35 (1995)
24. Caticha, A.: Consistency, amplitudes and probabilities in quantum theory. Phys. Rev. A **57**, 1572 (1998)
25. Hardy, L.: Quantum theory from five reasonable axioms. arXiv.org/quant-ph/0101012.
26. Brukner, Č., Zeilinger, A.: Information and fundamental elements of the structure of quantum theory. In: Castell, L., Ischebeck, O. (ed.) Time, Quantum, Information. Springer, Berlin (2003). arXiv:quant-ph/0212084
27. Brukner, Č., Zukowski, M.., Zeilinger, A.: The essence of entanglement. arXiv:quant-ph/0106119
28. Fuchs, C.: Quantum mechanics as quantum information (and only a little more). In: Khrenikov, A. (ed.) Quantum Theory: Reconstruction of Foundations. Växjo University Press, Växjo (2002). quant-ph/0205039
29. Clifton, R., Bub, J., Halvorson, H.: Characterizing quantum theory in terms of information theoretic constraints. Found. Phys. **33**, 1561 (2003)
30. Grangier, P.: Contextual objectivity: a realistic interpretation of quantum mechanics. Eur. J. Phys. **23**, 331 (2002). arXiv:quant-ph/0012122
31. Grangier, P.: Contextual objectivity and the quantum formalism, Proc. of the conference "Foundations of Quantum Information" (April 2004, Camerino, Italy). quant-ph/0407025
32. Luo, S.: Maximum Shannon entropy, minimum Fisher information, and an elementary game. Found. Phys. **32**, 1757 (2002)
33. Grinbaum, A.: Elements of information-theoretic derivation of the formalism of quantum theory. Int. J. Quant. Inf. **1**(3), 289–300 (2003)
34. Spekkens, R.: Evidence for the epistemic view of quantum states: a toy theory. Phys. Rev. A **75**, 032110 (2007)
35. Goyal, P.: Information-geometric reconstruction of quantum theory. Phys. Rev. A **78**, 052120 (2008)
36. D'Ariano, G.M.: Probabilistic theories: what is special about Quantum Mechanics? In: Philosophy of Quantum Information and Entanglement. Bokulich, A., Jaeger, G. (eds.), Cambridge University Press, Cambridge (to appear). arXiv:0807.4383
37. Dirac, P.A.M.: The Principles of Quantum Mechanics. Oxford University Press, Oxford (1982)
38. Aaronson, S.: Is quantum mechanics an island in theoryspace? In: Khrennikov, A. (ed.). Proc. of the Växjö Conference "Quantum Theory: Reconsideration of Foundations", 2004. quant-ph/0401062
39. Wootters, W.K.: Quantum mechanics without probability amplitudes. Found. Phys. **16**, 391 (1986)
40. Zyczkowski, K.: Quartic quantum theory: an extension of the standard quantum mechanics. J. Phys. A **41**, 355302–23 (2008)
41. Paterek, T., Dakic, B., Brukner, Č.: Theories of systems with limited information content. arXiv:0804.1423
42. Wootters, W.K., Fields, B.D.: Optimal state-determination by mutually unbiased measurements. Ann. Phys. **191**, 363 (1989)

43. Ivanovic, I.: Geometrical description of quantal state determination. J. Phys. A **14**, 3241 (1981)
44. Wootters, W.K., Zurek, W.H.: Complementarity in the double-slit experiment: quantum nonseparability and a quantitative statement of Bohr's principle. Phys. Rev. D **19**, 473 (1979)
45. Greenberger, D., Yasin, A.: Simultaneous wave and particle knowledge in a neutron interferometer. Phys. Lett. A **128**, 391 (1988)
46. Jaeger, G., Shimony, A., Vaidman, L.: Two interferometric complementarities. Phys. Rev. A **51**, 54–67 (1995)
47. Englert, B.G.: Fringe visibility and which-way information: an inequality. Phys. Rev. Lett. **77**, 2154 (1996)
48. Brukner, Č., Zeilinger, A.: Operationally invariant information in quantum measurements. Phys. Rev. Lett. **83**, 3354–3357 (1999)
49. Bohr, N.: In: Atomic Physics and Human Knowledge. Wiley, New York (1958)
50. Havrda, J., Charvát, F.: Quantification method of classification processes. Concept of structural a-entropy. Kybernetika **3**, 30–35 (1967)
51. Tsallis, C.: Possible generalizations of the Boltzmann-Gibbs statistics. J. Stat. Phys. **52**, 479–487 (1988)
52. Gisin, N.: Weinberg's non-linear quantum mechanics and superluminal communications. Phys. Lett. A **143**, 1–2 (1990)
53. Peres, A.: Nonlinear variants of Schrödinger's equation violate the second law of thermodynamics. Phys. Rev. Lett. **63**, 1114–1114 (1989)
54. Abrams, D.S., Lloyd, S.: Nonlinear quantum mechanics implies polynomial-time solution for NP-complete and ♯P problems. Phys. Rev. Lett. **81**, 3992–3995 (1998)
55. Hardy, L.: Probability theories in general and quantum theory in particular. Stud. Hist. Philos. Mod. Phys. **34**, 381–393 (2003)
56. Brukner, Č, Zeilinger, A.: Conceptual inadequacy of the Shannon information in quantum measurements. Phys. Rev. A **63**, 022113 (2001)
57. Nha, H., Zubairy, M.S.: Uncertainty inequalities as entanglement criteria for negative partial-transpose states. Phys. Rev. Lett. **101**, 130402 (2008)