Comparing the degrees of incompatibility inherent in probabilistic physical theories

Paul Busch, Teiko Heinosaari, Jussi Schultz and Neil Stevens

1 Department of Mathematics, University of York - York, YO10 5DD, UK, EU
2 Turku Centre for Quantum Physics, Department of Physics and Astronomy, University of Turku
FI-20014 Turku, Finland, EU
3 Dipartimento di Matematica, Politecnico di Milano - Piazza Leonardo da Vinci 32, I-20133 Milano, Italy, EU

received 2 May 2013; accepted in final form 19 June 2013
published online 18 July 2013

PACS 03.65.Ta – Foundations of quantum mechanics; measurement theory

Abstract – We introduce a new way of quantifying the degrees of incompatibility of two observables in a probabilistic physical theory and, based on this, a global measure of the degree of incompatibility inherent in such theories, across all observable pairs. This opens up a novel and flexible way of comparing probabilistic theories with respect to the nonclassical feature of incompatibility, raising many interesting questions, some of which will be answered here. We show that quantum theory contains observables that are as incompatible as any probabilistic physical theory can have if arbitrary pairs of observables are considered. If one adopts a more refined measure of the degree of incompatibility, for instance, by restricting the comparison to binary observables, it turns out that there are probabilistic theories whose inherent degrees of incompatibility are greater than that of quantum mechanics.

Copyright © EPLA, 2013

Quantum theory has a number of important features not known in classical physics, ranging from the superposition and indeterminacy principles formulated by the pioneers to the more recently discovered no-cloning and no-broadcasting theorems. It is an old problem to identify operationally significant properties of quantum theory that distinguish it from other probabilistic theories. In recent years many features have been under intensive investigation from this perspective, including information processing [1], optimal state discrimination [2], entropy [3], purification [4] and discord [5]. It has been found that some properties are quite generally valid in any non-classical (no-signaling) probabilistic theories while others are specifically quantum.

The existence of pairs of incompatible observables marks one of the most striking distinctions between quantum theory and classical physical theories. There are many manifestations of incompatibility, perhaps the most famous being the Heisenberg uncertainty principle [6]. However, there are many nonclassical probabilistic theories which also possess incompatible observables, and it will be of interest to compare quantum theory with alternative theories with respect to the feature of incompatibility.

To this end, we define the joint measurability region of any given pair of observables in a probabilistic theory. The joint measurability region describes the amount of added noise needed to make the observables jointly measurable. The global joint measurability feature of a probabilistic theory can then be characterized as the intersection of all the joint measurability regions associated with the theory.

We demonstrate that quantum theory contains observables that are as incompatible as observables in any probabilistic theory can be. Hence, we can say that, in a global sense, quantum theory has as great a degree of incompatibility as any other probabilistic theory. But if only binary observables are considered, the degree of incompatibility inherent in quantum theory is limited and we give an example of a probabilistic theory that contains maximally incompatible binary observables.

Our aim is thus to compare the incompatibility of pairs of observables in different probabilistic physical theories. We first need to set some minimal constraints.

A probabilistic theory is a framework that provides a description of physical systems in terms of states and observables with the following general properties:

i) The states of a system are represented by the elements of a convex subset of a real vector space.
An observable is represented as an affine mapping from the set of states into the set of probability distributions on some outcome space. For simplicity, we restrict ourselves here to observables with a finite or countable number of outcomes.

ii) Any affine mapping from the set of states into the set of probability distributions is a valid observable.

We consider a particular probabilistic theory (PT) as given by a family of convex sets of states with associated sets of observables that share some properties specific to that PT. One may think of each pair consisting of a set of states with associated set of observables as an instance of a PT representing a particular type of physical system.

Given a PT, we denote by $\mathcal{M}(j|\varrho)$ the probability of obtaining a measurement outcome $j$ when an observable $\mathcal{M}$ is measured in a state $\varrho$. Hence, $0 \leq \mathcal{M}(j|\varrho) \leq 1$ and $\sum_j \mathcal{M}(j|\varrho) = 1$. We will typically label the measurement outcomes by integers.

In quantum theory the states are described by density operators and observables correspond to positive-operator valued measures (POVMs) \cite{7}. Their duality is given by the trace formula (with $\varrho$ a density operator and $\mathcal{M}$ a POVM)

$$\mathcal{M}(j|\varrho) = \text{tr}[\varrho \mathcal{M}(j)].$$

(1)

Another example of a probabilistic theory is a classical theory, where the states are probability measures on a phase space $\Omega$ and observables are traditionally represented as functions $m: \Omega \rightarrow \mathbb{R}$; the associated affine maps from states $\varrho$ to probability distributions are then given by the formula

$$\mathcal{M}(j|\varrho) = \varrho(\{x \in m^{-1}(j)\}).$$

(2)

Continuing our discussion on general probabilistic theories, we note that it follows from the required properties i)–iii) that the set of observables is a convex set; a mixture of two observables is an observable. Physically mixing corresponds to an experiment where we switch between two measurement apparatuses with a random probability. We can directly write a mixture of two observables with the same set of measurement outcomes. If the sets of measurement outcomes differ, we can still write a mixture by first adding enough outcomes and then embedding both sets into $\Xi$.

Another consequence of the basic requirements is that every constant mapping $\varrho \mapsto p$, where $p$ is a fixed probability distribution, is an observable and we call it a trivial observable. A trivial observable $\Xi$ corresponds to a dice rolling experiment, where we randomly pick the measurement outcome according to a given fixed probability distribution, without manipulating the state at all. In quantum theory, trivial observables are described by POVMs $\mathcal{T}$ such that each operator $\mathcal{T}(j)$ is a multiple of the identity operator, i.e., $\mathcal{T}(j) = t_j I$ for some $0 \leq t_j \leq 1$ with $\sum_j t_j = 1$.

The concept of joint measurement can be defined in any probabilistic theory. Two observables $\mathcal{M}_1$ and $\mathcal{M}_2$ are jointly measurable if there exists an observable $\mathcal{M}$ such that

$$\sum_k \mathcal{M}(j,k|\varrho) = \mathcal{M}_1(j|\varrho), \quad \sum_j \mathcal{M}(j,k|\varrho) = \mathcal{M}_2(k|\varrho).$$

(3)

In this case $\mathcal{M}$ is called a joint observable of $\mathcal{M}_1$ and $\mathcal{M}_2$. If $\mathcal{M}_1$ and $\mathcal{M}_2$ are not jointly measurable, then we say that they are incompatible.

Any probabilistic theory contains jointly measurable pairs of observables. Namely, a trivial observable $\varrho \mapsto p$ is jointly measurable with any other observable; we can write a joint observable

$$\mathcal{M}(j,k|\varrho) = \mathcal{M}_1(j|\varrho)p(k)$$

(4)

for the trivial observable and any other observable $\mathcal{M}_1$. This simply corresponds to an experiment where we measure $\mathcal{M}_1$ and simultaneously roll a dice. It is a well known fact that, in quantum theory, an observable which is jointly measurable with all other observables is necessarily a trivial observable. Indeed, any POVM element of such an observable commutes with all projections and must therefore be a scalar multiple of the identity (e.g., \cite{8}, Theorem IV.1.3.1).

The following simple observation is a key ingredient for our discussion.

**Proposition 1.** Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be two observables and $0 \leq \lambda \leq 1$. Then $\lambda \mathcal{M}_1 + (1 - \lambda) \mathcal{M}_2$ and $(1 - \lambda)\mathcal{M}_1 + \lambda \mathcal{M}_2$ are jointly measurable for any choice of trivial observables $\mathcal{T}_1$ and $\mathcal{T}_2$.

This proposition can be proved with the following construction. First, let $p_1$ and $p_2$ be the probability distributions related to $\mathcal{T}_1$ and $\mathcal{T}_2$. We define an observable $\mathcal{M}$ by formula

$$\mathcal{M}(j,k|\varrho) = \lambda p_2(k) \mathcal{M}_1(j|\varrho) + (1 - \lambda)p_1(j) \mathcal{M}_2(k|\varrho).$$

(5)

For a fixed $\varrho$, the right-hand side is clearly a probability distribution. Moreover, the right-hand side is an affine mapping on $\varrho$; therefore $\mathcal{M}$ is an observable. The marginal observables are

$$\sum_k \mathcal{M}(j,k|\varrho) = \lambda \mathcal{M}_1(j|\varrho) + (1 - \lambda)p_1(j),$$

$$\sum_j \mathcal{M}(j,k|\varrho) = (1 - \lambda)\mathcal{M}_2(k|\varrho) + \lambda p_2(k).$$

This proves Proposition 1.

The physical idea behind this construction is the following. In each measurement run we flip a coin and, depending on the result, we measure either $\mathcal{M}_1$ or $\mathcal{M}_2$ in the input state $\varrho$. In this way we get a measurement outcome for either $\mathcal{M}_1$ or $\mathcal{M}_2$. In addition to this, we roll a dice and pretend that this is a measurement outcome for the other observable. In this way we get an outcome for both observables simultaneously. The overall observable is the one given in formula (5).

10002-p2
For two observables $\mathcal{M}_1$ and $\mathcal{M}_2$, we denote by $J(\mathcal{M}_1, \mathcal{M}_2)$ the set of all points $(\lambda, \mu) \in [0,1] \times [0,1]$ for which there exist trivial observables $\mathcal{E}_1, \mathcal{E}_2$ such that $\lambda \mathcal{M}_1 + (1-\lambda) \mathcal{E}_1$ and $\mu \mathcal{M}_2 + (1-\mu) \mathcal{E}_2$ are jointly measurable, and we call $J(\mathcal{M}_1, \mathcal{M}_2)$ the joint measurability region of $\mathcal{M}_1$ and $\mathcal{M}_2$. The joint measurability region thus characterizes how much noise (in terms of trivial observables) we need to add to obtain jointly measurable approximations of $\mathcal{M}_1$ and $\mathcal{M}_2$. Clearly, $\mathcal{M}_1$ and $\mathcal{M}_2$ are jointly measurable if and only if $(1,1) \in J(\mathcal{M}_1, \mathcal{M}_2)$.

The joint measurability region $J(\mathcal{M}_1, \mathcal{M}_2)$ is a convex region which can be plotted in the plane. To see this, let $(\lambda', \mu') \in J(\mathcal{M}_1, \mathcal{M}_2)$ and $(\lambda'', \mu'') \in J(\mathcal{M}_1, \mathcal{M}_2)$, then we have to show that $(\lambda, \mu) \in J(\mathcal{M}_1, \mathcal{M}_2)$ for $(\lambda, \mu) = t(\lambda', \mu') + (1-t)(\lambda'', \mu'')$. Thus let $\mathcal{M}'_1 = \lambda' \mathcal{M}_1 + (1-\lambda') \mathcal{E}_1$ and $\mathcal{M}'_2 = \mu' \mathcal{M}_2 + (1-\mu') \mathcal{E}_2$ be jointly measurable, and similarly for $\mathcal{M}''_1 = \lambda'' \mathcal{M}_1 + (1-\lambda'') \mathcal{E}_1$ and $\mathcal{M}''_2 = \mu'' \mathcal{M}_2 + (1-\mu'') \mathcal{E}_2$, with suitable choices of trivial observables. Then the observables $t \mathcal{M}'_1 + (1-t) \mathcal{M}''_1$ and $t \mathcal{M}'_2 + (1-t) \mathcal{M}''_2$ are jointly measurable ([9], Proposition 2).

Note that according to Proposition 1 the line $\{ (\lambda, (1-\lambda)) : 0 \leq \lambda \leq 1 \} \subseteq J(\mathcal{M}_1, \mathcal{M}_2)$. Moreover, it is trivially the case that $(0,0) \in J(\mathcal{M}_1, \mathcal{M}_2)$. The convexity of $J(\mathcal{M}_1, \mathcal{M}_2)$ then entails that the convex hull of the three points $(1,0)$, $(0,1)$ and $(0,0)$ is in $J(\mathcal{M}_1, \mathcal{M}_2)$, hence we have

$$\triangle \equiv \{ (\lambda, \mu) \in [0,1] \times [0,1] : \lambda + \mu \leq 1 \} \subseteq J(\mathcal{M}_1, \mathcal{M}_2).$$

As an example, suppose that we are within quantum theory and $\mathcal{M}_1$ and $\mathcal{M}_2$ correspond to spin-$\frac{1}{2}$ measurements in two orthogonal directions, say $x$ and $y$ axes. We then describe them with two POVMs $M_x$ and $M_y$, where

$$M_x(\pm 1) = \frac{1}{2} (I \pm \sigma_x), \quad M_y(\pm 1) = \frac{1}{2} (I \pm \sigma_y),$$

and $\sigma_x, \sigma_y$ are the usual Pauli matrices in $\mathbb{C}^2$. It has been shown in [10] that for the uniformly distributed trivial observable $\pm 1 \mapsto \frac{1}{2} I$ (hence describing an unbiased coin), the two observables $\lambda M_x + (1-\lambda) \frac{1}{2} I$ and $\mu M_y + (1-\mu) \frac{1}{2} I$ are jointly measurable if and only if $\lambda^2 + \mu^2 \leq 1$. It is also known ([9], Proposition 3) that this inequality is a necessary condition for the joint measurability of any pair $\lambda M_x + (1-\lambda) \mathcal{T}_1$ and $\mu M_y + (1-\mu) \mathcal{T}_2$, where $\mathcal{T}_1, \mathcal{T}_2$ are arbitrary trivial observables. Therefore, we conclude that

$$J(\mathcal{M}_x, \mathcal{M}_y) = \{ (\lambda, \mu) \in [0,1] \times [0,1] : \lambda^2 + \mu^2 \leq 1 \}. \quad (7)$$

This region is depicted in fig. 1.

![Fig. 1: (Color online) The region $J(\mathcal{M}_x, \mathcal{M}_y)$ for two orthogonal spin-$\frac{1}{2}$ measurements is a quadrant of the unit disk. The region $\triangle$ (light) is a subset of the joint measurability region $J_{PT}$ (colored) for any probabilistic theory. The larger the surplus region (dark) is, the more jointly measurable the theory globally is. If (a) and (b) are joint measurability regions for two different probabilistic theories, then we can conclude that (b) represents a greater degree of incompatibility than (a).](image)

In addition to describing the incompatibility of pairs of observables, the concept of a joint measurability region also provides a means to compare the degrees of incompatibility inherent in entire theories. A global joint measurability feature of a probabilistic theory $PT$ is characterized by the intersection of all the sets $J(\mathcal{M}_1, \mathcal{M}_2)$ across all instances of $PT$, and we denote

$$J_{PT} = \{ (\lambda, \mu) \in [0,1] \times [0,1] : (\lambda, \mu) \in J(\mathcal{M}_1, \mathcal{M}_2) \}$$

for all pairs of observables $\mathcal{M}_1$ and $\mathcal{M}_2$ in all instances of $PT$.

We call $J_{PT}$ the joint measurability region for $PT$. We always have $\triangle \subseteq J_{PT}$, but $J_{PT}$ can be larger than $\triangle$. The larger the surplus region is, the more jointly measurable the theory globally is. If (a) and (b) are joint measurability regions for any pair $\mathcal{M}_1, \mathcal{M}_2$, while the surplus region (dark) depends on the specific pair under consideration.

![Fig. 2: (Color online) The region $\triangle$ (light) is a subset of the joint measurability region $J_{PT}$ (colored) for any probabilistic theory. The larger the surplus region (dark) is, the more jointly measurable the theory globally is. If (a) and (b) are joint measurability regions for two different probabilistic theories, then we can conclude that (b) represents a greater degree of incompatibility than (a).](image)
In quantum theory every observable \( \mathfrak{M} \) corresponds to a unique POVM \( \mathbf{M} \) by equation (1). We will prove that for any pair \( (\lambda, \mu) \notin \Delta \), there are quantum observables \( M_1 \) and \( M_2 \) such that the mixtures \( \lambda M_1 + (1 - \lambda)T_1 \) and \( \mu M_2 + (1 - \mu)T_2 \) are incompatible with any choice of trivial observables \( T_1, T_2 \). Our proof is based on a recent result [11] on the joint measurability region for two complementary observables, which is a generalization of the result illustrated in fig. 1.

Proof of Theorem 1. We have earlier seen that \( \Delta \subseteq J_{\text{PT}} \), so we need to show that \( J_{\text{QT}} \subseteq \Delta \). Let \( (\lambda, \mu) \notin \Delta \), i.e., \( \lambda + \mu > 1 \). Fix \( \epsilon > 0 \) such that \( \lambda + \mu > 1 + \epsilon \). We then choose \( d \) to be a positive integer satisfying

\[
\frac{\sqrt{d} - 1}{d - 1} \leq \epsilon.
\]  

(8)

(This can be done since the left hand side \( \to 0 \) when \( d \to \infty \).) We will consider a quantum system that is described by a \( d \)-dimensional Hilbert space \( \mathcal{H}_d \). Let \( \{|\varphi_j\rangle\}_{j=0}^{d-1} \) be an orthonormal basis for \( \mathcal{H}_d \). We define another orthonormal basis \( \{\psi_k\}_{k=0}^{d-1} \) for \( \mathcal{H}_d \) by

\[
\psi_k = 1/\sqrt{d} \sum_j e^{-2\pi i jk/d} \varphi_j.
\]  

(9)

The orthonormal bases \( \{\varphi_j\}_{j=0}^{d-1} \) and \( \{\psi_k\}_{k=0}^{d-1} \) are mutually unbiased, i.e., \( \langle \varphi_j | \psi_k \rangle = \text{const} \). We define two POVMs \( M_1 \) and \( M_2 \) by

\[
M_1(j) = |\varphi_j\rangle \langle \varphi_j|, \quad M_2(k) = |\psi_k\rangle \langle \psi_k|.
\]  

(10)

We thus obtain a pair of \( d \)-outcome observables on \( \mathcal{H}_d \). Since \( M_1 \) and \( M_2 \) consist of projections and \( M_1(j)M_2(k) \neq M_2(k)M_1(j) \), it follows that they are incompatible.

As proved in [11], the observables \( \lambda' M_1 + (1 - \lambda')T_1 \) and \( \mu' M_2 + (1 - \mu')T_2 \) are incompatible for any choice of trivial observables \( T_1, T_2 \) whenever

\[
\lambda' + \mu' > 1 + \frac{\sqrt{d} - 1}{d - 1}.
\]  

(11)

Since

\[
\lambda + \mu > 1 + \epsilon > 1 + \frac{\sqrt{d} - 1}{d - 1},
\]  

(12)

we conclude that \( (\lambda, \mu) \notin J_{\text{QT}} \).

Using the ideas of the proof of Theorem 1, we can also show that the conclusion \( J_{\text{QT}} = \Delta \) can be reached by using a single pair of incompatible observables if we consider an infinite dimensional system and observables with a countably infinite number of outcomes.

Let \( \mathcal{H} \) be an infinite dimensional Hilbert space and write it as a direct sum of finite \( d \)-dimensional Hilbert spaces \( \mathcal{H}_d \), \( \mathcal{H} = \bigoplus_{d=2}^\infty \mathcal{H}_d \). In each \( \mathcal{H}_d \) consider a pair of mutually unbiased orthonormal bases \( \{|\varphi_j^d\rangle\}_{j=0}^{d-1} \) and \( \{|\psi_k^d\rangle\}_{k=0}^{d-1} \), where the latter is obtained from the first by the formula (9). We define two POVMs \( N_1 \) and \( N_2 \) via

\[
N_1(d, j) = |\varphi_j^d\rangle \langle \varphi_j^d|, \quad N_2(d, k) = |\psi_k^d\rangle \langle \psi_k^d|.
\]  

(13)

These observables act in the infinite dimensional Hilbert space \( \mathcal{H} \) and \( d \) in (13) is an index labeling the different outcomes. The outcome space of \( N_1 \) and \( N_2 \) is \( \Omega_\infty \equiv \{(d, j): d = 2, 3, \ldots, j = 0, \ldots, d - 1\} \).

Theorem 2. The observables \( N_1 \) and \( N_2 \) defined in (13) satisfy \( J(N_1, N_2) = \Delta \).

Proof of Theorem 2. Let \( p_1 \) and \( p_2 \) be two probability distributions defined on \( \Omega_\infty \). Assume that \( \lambda + \mu > 1 \) and define two observables \( N_{1,\lambda}, N_{2,\mu} \) via

\[
N_{1,\lambda}(d, j) = \lambda |\varphi_j^d\rangle \langle \varphi_j^d| + (1 - \lambda) p_1(d, j) I, \quad N_{2,\mu}(d, k) = \mu |\psi_k^d\rangle \langle \psi_k^d| + (1 - \mu) p_2(d, k) I.
\]  

(14)

We need to show that \( N_{1,\lambda} \) and \( N_{2,\mu} \) are incompatible. To prove this, we make the counter assumption that \( N_{1,\lambda}, N_{2,\mu} \) are jointly measurable. This implies that for any projection \( P \) on \( \mathcal{H} \), the projected observables \( PN_{1,\lambda}P \) and \( PN_{2,\mu}P \) acting on a subspace \( P \mathcal{H} \) are jointly measurable. (If \( G \) is a joint observable of two observables \( M_1, M_2 \), then \( PG \) is a joint observable of \( PM_1P, PM_2P \) in \( P \mathcal{H} \).) Especially, the projections of \( N_{1,\lambda} \) and \( N_{2,\mu} \) to any subspace \( \mathcal{H}_d \) should be jointly measurable. But from the result cited in the proof of Theorem 1 we know that for \( d \) large enough, the projections to \( \mathcal{H}_d \) are incompatible. Hence, \( N_{1,\lambda} \) and \( N_{2,\mu} \) are incompatible.

We note that the observables \( N_1 \) and \( N_2 \) defined in (13) are not the only pair satisfying \( J(N_1, N_2) = \Delta \). Namely, we can modify \( N_1 \) and \( N_2 \) in any chosen subspace \( \mathcal{H}_d \) but the conclusion \( J(N_1, N_2) = \Delta \) is still true since it depends on the fact that \( N_1 \) and \( N_2 \) contain mutually unbiased bases in arbitrarily high dimension.

An interesting problem within quantum theory would be to try to find a characterization of all pairs of quantum observables \( M_1, M_2 \) that satisfy \( J(M_1, M_2) = \Delta \). In particular, we may ask if maximally incompatible observables can exist in a finite dimensional Hilbert space, or if they can have a finite number of outcomes. Since two mutually unbiased bases are expected to be among the most incompatible observable pairs in a fixed dimension \( d \), our construction in the proof of Theorem 1 suggests that the answer to the first question would be negative. A proof of this claim is, however, lacking.

As for the second question, we can present a partial answer by investigating the joint measurability region in the case of pairs of binary quantum observable. Our aim is to show that

\[
\{(\lambda, \mu) \in [0, 1] \times [0, 1]: \lambda^2 + \mu^2 \leq 1\} \subseteq J(M_1, M_2)
\]

for any binary observables \( M_1 \) and \( M_2 \), regardless of the dimension of the Hilbert space. In other words, we will
show that two orthogonal spin observables are as incompatible as any binary observables can be.

To this end, let us note that two binary quantum observables are incompatible if and only if they enable a violation of the Bell-CHSH inequality [12]. We must therefore look at the Bell expression

\[ B = |\langle M_1 N_1 \rangle + \langle M_2 N_2 \rangle + \langle M_2 N_1 \rangle - \langle M_2 N_2 \rangle|. \]

Let us denote \( \alpha = \langle M_1 N_1 \rangle + \langle M_2 N_2 \rangle \) and \( \beta = \langle M_2 N_1 \rangle - \langle M_2 N_2 \rangle \). By [13], Theorem 1, there exist unit vectors \( x_j, y_1, y_2 \in \mathbb{R}^4 \) such that \( \langle M_j N_k \rangle = x_j \cdot y_k \) for \( j, k = 1, 2 \); and conversely, given any quadruple of unit vectors there exist a corresponding set of binary observables and a bipartite state such that this equality holds. In particular, we have \( \alpha = x_1 \cdot (y_1 + y_2) \) and \( \beta = x_2 \cdot (y_1 - y_2) \) so that an application of the Cauchy-Schwarz inequality along with the parallelogram law yields

\[ \alpha^2 + \beta^2 \leq \|x_1\|^2 \|y_1 + y_2\|^2 + \|x_2\|^2 \|y_1 - y_2\|^2 = 2 \|y_1\|^2 + 2 \|y_2\|^2 = 4. \]

By choosing the unit vectors appropriately we also see that any pair \( (\alpha, \beta) \) satisfying this condition can be obtained.

If we now mix the observables \( M_j \) with the trivial observable \( T(z) = \frac{1}{2}I \) with some weights \( \lambda \) and \( \mu \), we see that the pair \( (\alpha, \beta) \) turns into \( (\lambda \alpha, \mu \beta) \), thus changing the Bell expression from \( |\alpha + \beta| \) to \( |\lambda \alpha + \mu \beta| \). We must therefore determine those \( (\lambda, \mu) \) for which \( |\lambda \alpha + \mu \beta| \leq 2 \) for all \( (\alpha, \beta) \) satisfying \( \alpha^2 + \beta^2 \leq 4 \) (see fig. 3). But the boundary curve for this region is obtained when the equations \( (\alpha/\lambda)^2 + (\beta/\mu)^2 = 4 \) and \( \alpha + \beta = 2 \) have at most one common solution. By inserting \( \beta = 2 - \alpha \) into the first equation the problem reduces to determining when the discriminant is negative or zero, and one readily verifies that this is the case exactly when \( \lambda^2 + \mu^2 \leq 1 \).

In conclusion, given any pair of binary observables \( M_1 \) and \( M_2 \), and weights \( \lambda \) and \( \mu \) with \( \lambda^2 + \mu^2 \leq 1 \), the mixtures \( \lambda M_1 + (1 - \lambda)T \) and \( \mu M_2 + (1 - \mu)T \) cannot be used to violate the Bell-CSHS inequality and must therefore be jointly measurable. We note that in the case \( \mu = \lambda \) the same result using a different technique has been obtained by Banik et al. [14].

Although Theorem 2 shows that quantum theory contains pairs of observables that are maximally incompatible, the strictly larger joint measurability region when restricting to binary observables suggests that more fine-grained quantifications of the global degree of incompatibility between observables might not rank quantum theory among the most extreme theories in this respect. The example below will show that when restricting to just binary observables, it is indeed possible for a theory to have the smallest possible joint measurability region. In that sense such a theory must be considered to embody a strictly greater degree of incompatibility than quantum theory.

Consider any probabilistic theory, which contains a state space isomorphic to a square, by which we mean the convex hull of four different points \( s_1, s_2, s_3, s_4 \) in \( \mathbb{R}^2 \) satisfying \( s_1 + s_4 = s_2 + s_3 \), for instance \( s_1 = (0, 0), s_2 = (0, 1), s_3 = (1, 0) \) and \( s_4 = (1, 1) \). We will show that there is a pair of binary observables which are maximally incompatible. Let \( M_1 \) and \( M_2 \) be binary observables that pick out the right and top sides of the square, respectively, i.e.,

\[
\begin{align*}
&M_1(+|s_1) = M_1(+|s_2) = 0, \\
&M_1(+|s_3) = M_1(+|s_4) = 1, \\
&M_2(+|s_1) = M_2(+|s_3) = 0, \\
&M_2(+|s_2) = M_2(+|s_4) = 1.
\end{align*}
\]

Proposition 2. For the binary observables \( M_1 \) and \( M_2 \) defined in (15), \( J(M_1, M_2) = \Delta \).

Proof. Suppose that there exists a joint observable \( M \) for \( \lambda M_1 + (1 - \lambda)T_1 \) and \( \mu M_2 + (1 - \mu)T_2 \) where \( T_1 \) and \( T_2 \) are trivial observables. Let \( p_1 \) and \( p_2 \) be the probability distributions associated to \( T_1 \) and \( T_2 \) so that we have for any state \( \rho \)

\[
\begin{align*}
&M(+, +|\rho) + M(+, -|\rho) = \lambda M_1(+|\rho) + (1 - \lambda)p_1(+), \\
&M(-, +|\rho) + M(-, -|\rho) = \lambda M_1(-|\rho) + (1 - \lambda)p_1(-), \\
&M(+, +|\rho) + M(-, -|\rho) = \mu M_2(+|\rho) + (1 - \mu)p_2(+), \\
&M(-, +|\rho) + M(-, -|\rho) = \mu M_2(-|\rho) + (1 - \mu)p_2(-).
\end{align*}
\]

Any \( M \) satisfying such marginal properties will be correctly normalised, but to be a valid observable, all the components of \( M \) must take positive values on the points \( s_i \). In particular, we must have

\[
\begin{align*}
&M(+, -|s_2) = (1 - \lambda)p_1(+) - M(+, +|s_2) \geq 0, \\
&M(-, +|s_3) = (1 - \mu)p_2(+) - M(-, +|s_3) \geq 0, \\
&M(-, -|s_4) = 1 + M(+, +|s_4) - \lambda - (1 - \lambda)p_1(+) - \mu - (1 - \mu)p_2(+) \geq 0.
\end{align*}
\]
Rewriting the last of these inequalities and invoking the defining property on the $s_i$ gives
\[
\lambda + \mu \leq 1 - (1 - \lambda)p_1(+)*(1 - \mu)p_2(+) + \mathcal{M}(+, +|s_i)
\]
\[
\leq - (1 - \lambda)p_1(+) + \mathcal{M}(+, +|s_2)
\]
\[
- (1 - \mu)p_2(+) + \mathcal{M}(+, +|s_3)
\]
\[
+ 1 - \mathcal{M}(+, +|s_1) \leq 1,
\]
where the final step comes about from invoking the positivity of $\mathcal{M}$ on $s_1$.

The result of Proposition 2 does not come as a surprise in light of the fact that the barrier to maximal incompatibility of binary quantum observables comes from the connection with a Bell-CHSH inequality. Indeed, square-shaped state spaces have been used in a model of a probabilistic theory containing the PR boxes which violate such an inequality to its maximal possible value.

We note that the conclusion of Proposition 2 is not restricted to the square state space. Consider any state space containing a square whose vertices $s_i$ are extreme points of the state space and whose boundary lines lie on the boundary of the state space; assume further that opposite sides of the square are contained in parallel hyperplanes that do not intersect with the interior of the state space. These two pairs of hyperplanes define effects whose values on the $s_i$ satisfy eq. (15). It follows that the proof of Proposition 2 can be adopted in such cases. Examples are given by state space of the following shapes: pyramid, double pyramid, cube, cylinder.

The fact that the restriction to just binary observables allows one to differentiate between probabilistic theories that both contain maximally incompatible observables suggests that a more fine-grained global measure of the degree of incompatibility is needed if the aim is to pick out a single theory as the one containing overall the most incompatible pairs of observables. For instance, for a given probabilistic theory PT we may define $J_{PT}^{(d)}$ to be the joint measurability region for all possible $d$-outcome observables in PT. Since increasing the number of outcomes of observables by simply adding outcomes that never occur does not change the properties of incompatibility, we immediately have $J_{PT}^{(d+1)} \subseteq J_{PT}^{(d)}$. By comparing the regions in different theories for different values of $d$ we obtain a more fine-grained way of comparing the degrees of incompatibility within the theories. It may even turn out that in this sense quantum theory embodies globally the least amount of incompatibility among the theories containing maximally incompatible observables. However, this is still an open question and a topic for future investigations.

** * **

The authors wish to thank Tom Bullock for comments on an earlier version of this paper. TH and JS acknowledge financial support from the Academy of Finland (grant No. 138135). JS also acknowledges financial support from the Italian Ministry of Education, University and Research (FIRB project RBFR10COAQ). NS gratefully acknowledges support through the award of an Annie Currie Williamson PhD Bursary at the University of York. Finally, the authors wish to thank an anonymous referee for suggesting a number of improvements to the presentation of the material.

REFERENCES

[1] Barrett J., Phys. Rev. A, 75 (2007) 032304.
[2] Kimura G., Miyadera T. and Imai H., Phys. Rev. A, 79 (2009) 062306.
[3] Barnum H., Barrett J., Orloff Clark L., Leifer M., Spekkens R., Stepanik N., Wilce A. and Wilke R., New J. Phys., 12 (2010) 033024.
[4] Chiribella G., D’Ariano G. M. and Perinotti P., Phys. Rev. A, 84 (2011) 012311.
[5] Perinotti P., Phys. Rev. Lett., 108 (2012) 120502.
[6] Busch P., Heinonen T. and Lahti P., Phys. Rep., 452 (2007) 155.
[7] Busch P., Grabowski M. and Lahti P. J., Operational Quantum Physics (Springer-Verlag, Berlin) 1997.
[8] Ludwig G., Foundations of Quantum Mechanics I (Springer-Verlag, Berlin) 1983.
[9] Busch P. and Heinosaari T., Quantum Inf. Comput., 8 (2008) 0797.
[10] Busch P., Phys. Rev. D, 33 (1986) 2253.
[11] Carmeli C., Heinosaari T., Toigo A., Phys. Rev. A, 85 (2012) 012109.
[12] Wolf M. M., Perez-Garcia D. and Fernandez C., Phys. Rev. Lett., 103 (2009) 230402.
[13] Ciruelson B. S., Lett. Math. Phys., 4 (1980) 93.
[14] Banik M., Gazi Md. R., Ghosh S. and Kar G., Phys. Rev. A, 87 (2013) 052125.

Paul Busch et al.