Observers of quantum systems cannot agree to disagree

Patricia Contreras-Tejada [1], Giannicola Scarpa [2✉], Aleksander M. Kubicki [3], Adam Brandenburger[4] & Pierfrancesco La Mura[5]

Is the world quantum? An active research line in quantum foundations is devoted to exploring what constraints can rule out the postquantum theories that are consistent with experimentally observed results. We explore this question in the context of epistemics, and ask whether agreement between observers can serve as a physical principle that must hold for any theory of the world. Aumann's seminal Agreement Theorem states that two observers (of classical systems) cannot agree to disagree. We propose an extension of this theorem to no-signaling settings. In particular, we establish an Agreement Theorem for observers of quantum systems, while we construct examples of (postquantum) no-signaling boxes where observers can agree to disagree. The PR box is an extremal instance of this phenomenon. These results make it plausible that agreement between observers might be a physical principle, while they also establish links between the fields of epistemics and quantum information that seem worthy of further exploration.

[1] Instituto de Ciencias Matemáticas, 28049 Madrid, Spain. [2] Escuela Técnica Superior de Ingeniería de Sistemas Informáticos, Universidad Politécnica de Madrid, 28031 Madrid, Spain. [3] Departamento de Análisis Matemático y Matemática Aplicada, Universidad Complutense de Madrid, 28040 Madrid, Spain. [4] Stern School of Business, Tandon School of Engineering, NYU Shanghai, New York University, New York, NY 10012, USA. [5] HHL Leipzig Graduate School of Management, 04109 Leipzig, Germany. ✉email: g.scarpa@upm.es

Quantum mechanics famously made its creators uncomfortable. It is highly counterintuitive and, almost a century after its introduction, it still sparks much conceptual and philosophical discussion. Indeed, an active line of research in quantum foundations deals with the problem of singling out quantum theory from other post-classical physical theories. This field is a delicate balance between proposals for new theories that are 'tidier' than quantum mechanics[1,2] and proposals for desirable physical principles that such theories should obey[3–7].

In the domain of classical probability theory, Aumann proved that Bayesian agents cannot agree to disagree[8]. A slightly more general restatement of Aumann's theorem, which we will refer to as the classical agreement theorem, states that, if Alice and Bob, based on their partial information, assign probabilities q_A, q_B, respectively, to perfectly correlated events, and these probabilities are common certainty between them, then $q_A = q_B$. "Certainty" means assigning probability 1, and "common certainty" means that Alice is certain about q_B; Bob is certain about q_A; Alice is certain about Bob being certain about q_A; Bob is certain about Alice being certain about q_A; and so on infinitely.

This result is considered a basic requirement in classical epistemics, which is the formal study of the knowledge and beliefs of the agents in a system. The classical agreement theorem has been used to show that two risk-neutral agents, starting from a common prior, cannot agree to bet with each other[9], to prove "no-trade" theorems for efficient markets[10], and to establish epistemic conditions for Nash equilibrium[11].

Focusing on the quantum domain, a fundamental result of quantum mechanics is that no local hidden-variable theory can model the results of all quantum experiments[12]. This implies that the classical Bayesian model does not apply, so the classical agreement theorem need not hold. The question then arises: Can observers of quantum mechanical phenomena agree to disagree?

In this work, we answer the above question in the negative. We define two notions of disagreement inspired by Aumann's theorem. One is a direct analogue to the classical agreement theorem, and the other one relaxes the common certainty condition while requiring that the probability estimates differ maximally. We find that neither kind of disagreement occurs for classical or quantum systems. However, both kinds of disagreement do occur in postquantum environments. In fact, we characterize no-signaling distributions displaying these behaviors. We then put our two characterizations together and search for distributions that satisfy both notions of disagreement: We find that the PR box[3] is of this kind—i.e., it displays extremal disagreement in the above sense. Since the PR box is also an extreme instance of a no-signaling box as a non-local resource[3,13], our findings suggest a deeper relation between the quantification of disagreement and the quantification of non-locality.

If a physical theory were to allow agents to agree to disagree, then undesirable consequences in the settings of refs. [9–11] could happen. This is why the impossibility of agreeing to disagree is a desirable feature for all physical theories, and why we propose that it should be elevated to a physical principle. Its simplicity makes it convenient for testing the consistency of new postquantum theories.

Results

Classical agreement theorem. We start with an intuition about the setup behind the classical agreement theorem. Suppose that Alice and Bob share a classical system, which can, thus, be described by a local hidden-variable model (in Aumann's language, each value of the variable represents a state of the world). But the observers do not know which value of the hidden variable is the one that holds (i.e., which is the true state of the world).

Instead, each observer can perform only one local measurement on the system. Each measurement corresponds to a partition over the values of the variable, and the result reveals which partition element contains the value that holds. The probability of each outcome is the sum of the probabilities of the values in the corresponding partition element.

Suppose, also, that Alice is interested in estimating the probability of an event (i.e., a set of values of the variable) that does not correspond to an element in her partition. Then, she can calculate only the conditional probability of the event given the outcome of her local measurement by Bayesian inference. The same applies for Bob. Suppose that Alice and Bob are interested in events that are perfectly correlated (i.e., with probability 1, either both happen or neither happens). Then, the classical agreement theorem says that, if their estimates are common certainty, they must be equal.

Common certainty means that Alice is certain of (i.e., assigns probability 1 to) Bob's estimate, Bob is certain of Alice's estimate, Alice is certain of Bob's certainty of Alice's estimate, and so on.

When formalizing these notions, we refer to a probability space, together with some given partitions, as a (classical) ontological model. Ontological models appearing in the literature (see, e.g., ref. [14]) also contain a set of preparations underlying the distribution over the probability space, and the partitions are usually phrased in terms of measurements and outcomes. However, we consider preparations implicit and use the language of partitions to bridge the gap between classical probability spaces and no-signaling boxes more smoothly.

For the sake of simplicity and following Aumann, we restrict our analysis to two observers, Alice and Bob. Aumann's original theorem considers common knowledge of one single event of interest to both observers. We provide a slight generalization with common certainty about two perfectly correlated events of interest, one for each observer. This allows us to move to the framework of no-signaling boxes that we will use later. This is what we call the classical agreement theorem. This terminology can be further motivated by the fact that, for purely classical situations, both statements—the original Aumann's theorem and our formulation with perfectly correlated events—can be proven to be equivalent (as long as states of the world with null probability are ignored, as in ref. [15]).

Consider a probability space $(\Omega, \mathcal{E}, \mathsf{P})$ where Ω is a finite set of possible states of the world; $\mathcal{E}$ is its power set (i.e., the set of events); and P is a probability measure over Ω. We will consider two events $E_A, E_B \in \mathcal{E}$ of interest to Bob and Alice, respectively (the choice of subscripts will become clear later). We will assume that they are perfectly correlated: $\mathsf{P}(E_A \backslash E_B) = \mathsf{P}(E_B \backslash E_A) = 0$.

Fix partitions $\mathcal{P}_A, \mathcal{P}_B$ of Ω for Alice and Bob, respectively. For convenience, assume that all members of the join (coarsest common refinement) of $\mathcal{P}_A$ and $\mathcal{P}_B$ are non-null. For a state $\omega \in \Omega$, $\mathcal{P}_A(\omega)$ ($\mathcal{P}_B(\omega)$) is the partition element of Alice's (Bob's) that contains ω. For each $n \in \mathbb{N}$, fix numbers $q_A, q_B \in [0, 1]$ and consider the following sets:

$$A_0 = \left\{ \omega \in \Omega : \mathsf{P}(E_B | \mathcal{P}_A(\omega)) = q_A \right\}, \tag{1}$$

$$B_0 = \left\{ \omega \in \Omega : \mathsf{P}(E_A | \mathcal{P}_B(\omega)) = q_B \right\}, \tag{2}$$

$$A_{n+1} = \left\{ \omega \in A_n : \mathsf{P}(B_n | \mathcal{P}_A(\omega)) = 1 \right\}, \tag{3}$$

$$B_{n+1} = \left\{ \omega \in B_n : \mathsf{P}(A_n | \mathcal{P}_B(\omega)) = 1 \right\}. \tag{4}$$

Here, the set A_0 is the set of states ω such that Alice assigns probability q_A to event E_B; the set B_1 is the set of states ω such that Bob assigns probability q_B to event E_A and probability 1 to the states in A_0—i.e., states where Bob assigns probability q_B to E_A

and is certain that Alice assigned probability q_A to E_B; and so on, and similarly for B_0, A_1, etc.

If a state of the world ω^* is in sets A_n and B_n, for all $n \in \mathbb{N}$, then, at ω^*, Alice assigns probability q_A to E_B, Bob is certain that Alice assigns probability q_A to E_B, and Alice is certain that Bob is certain that Alice assigns probability q_A to E_B, and Alice is certain that Bob is certain that... and so on indefinitely, and also vice versa about Bob assigning probability q_B to E_A. In sum, there is common certainty at ω^* that Alice assigns probability q_A to E_B and that Bob assigns probability q_B to E_A.

More formally, it is common certainty at a state $\omega^* \in \Omega$ that Alice assigns probability q_A to E_B and that Bob assigns probability q_B to E_A if

$$\omega^* \in A_n \cap B_n \quad \forall n \in \mathbb{N}. \tag{5}$$

If equation (5) does not hold for all $n \in \mathbb{N}$, but, instead, only for $n \le N$ for a certain $N \in \mathbb{N}$, then we talk about Nth-order mutual certainty.

We now state the classical agreement theorem that will be the basis of our work. All proofs are given in the Supplementary Material.

Theorem 1. Fix a probability space $(\Omega, \mathcal{E}, \mathsf{P})$, where E_A and E_B are perfectly correlated events. If it is common certainty at a state $\omega^* \in \Omega$ that Alice assigns probability q_A to E_B and Bob assigns probability q_B to E_A, then $q_A = q_B$.

Defining agreement for observers of no-signaling systems. We now map the classical agreement theorem into the no-signaling framework, in order to explore its applicability beyond the classical realm.

We consider no-signaling distributions, or boxes[3], of the form

$$\left\{ p(ab|xy) \right\}_{a \in \mathcal{A}, b \in \mathcal{B}, x \in \mathcal{X}, y \in \mathcal{Y}}, \tag{6}$$

where x, y and a, b are Alice's and Bob's input and output, respectively, and $\mathcal{A}, \mathcal{B}, \mathcal{X}, \mathcal{Y}$ are index sets, not necessarily of the same size, and which satisfy

$$\sum_a p(ab|xy) = \sum_a p(ab|x'y), \tag{7}$$

$$\sum_b p(ab|xy) = \sum_b p(ab|xy'), \tag{8}$$

for all x, x', y, y'.

A no-signaling box is local if there exist probability distributions $\{p_\lambda : \lambda \in \Lambda\}$, $\{p_A(a|x\lambda) : (a, x, \lambda) \in \mathcal{A} \times \mathcal{X} \times \Lambda\}$, $\{p_B(b|y\lambda) : (b, y, \lambda) \in \mathcal{B} \times \mathcal{Y} \times \Lambda\}$, such that

$$p(ab|xy) = \sum_{\lambda \in \Lambda} p_\lambda \, p_A(a|x\lambda) \, p_B(b|y\lambda), \tag{9}$$

for each a, b, x, y, where Λ is an index set. It is quantum if, for each x, y, there exist POVMs $\{E_x^a\}_{a \in \mathcal{A}}$, $\{F_y^b\}_{b \in \mathcal{B}}$ and a quantum state ρ such that

$$p(ab|xy) = \mathrm{tr}\left(E_x^a \otimes F_y^b \rho \right), \tag{10}$$

for each a, b. The set of local boxes is strictly included in the set of quantum boxes, which is, in turn, strictly included in the set of no-signaling boxes. No-signaling boxes that are not quantum are termed postquantum.

We now show that we can associate a no-signaling box with any ontological model, and vice versa. Let $\mathcal{A}, \mathcal{B}, \mathcal{X}, \mathcal{Y}$ be index sets. Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space, and, for each $x \in \mathcal{X}$, let $\{\mathsf{A}_x^a : a \in \mathcal{A}\}$ be a partition of the states $\omega \in \Omega$ where $a \in \mathcal{A}$ denotes the partition elements. Similarly, for each $y \in \mathcal{Y}$, let $\{\mathsf{B}_y^b : b \in \mathcal{B}\}$ be another partition of the states $\omega \in \Omega$, where $b \in \mathcal{B}$

denotes the partition elements. According to that, we can understand labels $x \in \mathcal{X}$, $y \in \mathcal{Y}$ as inputs—this information fixes what partition Alice and Bob look at—and $a \in \mathcal{A}$, $b \in \mathcal{B}$ as outputs—this is the information that the observers gain by observing their corresponding partitions. This terminology will shortly become very natural.

With all the above, $\{(\Omega, \mathcal{F}, \mathsf{P}), \{\mathsf{A}_x^a, \mathsf{B}_y^b\}_{a,b,x,y}\}$ is an ontological model that we now want to associate to a no-signaling box that reproduces its statistics. In this ontological model, given inputs $x \in \mathcal{X}$, $y \in \mathcal{Y}$, the probability of obtaining outputs $a \in \mathcal{A}$, $b \in \mathcal{B}$ is given by $\mathsf{P}(\mathsf{A}_x^a \cap \mathsf{B}_y^b)$. This simple observation leads us to construct the no-signaling box

$$p(a, b|x, y) := \mathsf{P}\left(\mathsf{A}_x^a \cap \mathsf{B}_y^b \right), \qquad \forall (a, b, x, y) \in \mathcal{A} \times \mathcal{B} \times \mathcal{X} \times \mathcal{Y}. \tag{11}$$

It can be verified that the probabilities p are non-negative, normalized, and no-signaling.

The converse process of finding an ontological model starting from a no-signaling box can be also performed, as we show in Supplementary Note 2. Remarkably, this can be accomplished even in the case in which the no-signaling box is non-local, obtaining an ontological model with a quasi-probability measure (i.e., one which allows for negative values, which still sum to 1) instead of standard positive probabilities[16,17]. (The appearance of quasi-probabilities here should not surprise the reader. In fact, one cannot hope to obtain ontological models with only non-negative probabilities for post-classical no-signaling boxes, since this would provide local hidden-variable models that contradict, for instance, Bell's theorem. In any case, the use of this mathematical tool has been well rooted in the study of quantum mechanics since its origins—see ref. [14] for a nice review of this subject.) This makes it possible to translate results from one framework to the other, something that might be of interest in order to establish further connections between epistemics and quantum theory. However, from now on, we focus on no-signaling boxes and leave this digression aside in the rest of the main text.

With the association between ontological models and no-signaling boxes in mind, we next define common certainty of disagreement for no-signaling boxes. The idea is to reinterpret the definitions in Section 2.1 in this latter setting.

We first propose a meaning for the events of interest (previously identified as E_A, E_B) in the present setting. Now, these events correspond to some set of outcomes, given that the no-signaling box was queried with some particular inputs. For the sake of concreteness, we fix these inputs to be $x = 1$, $y = 1$ and the outcomes of interest to be $a = 1$, $b = 1$. This motivates us to consider the events $F_A = \{(1, b, 1, y) : b \in \mathcal{B}, y \in \mathcal{Y}\}$ (on Alice's side) and $F_B = \{(a, 1, x, 1) : a \in \mathcal{A}, x \in \mathcal{X}\}$ (on Bob's side). Then, we say that F_A and F_B are perfectly correlated when

$$p(a, b|x = 1, y = 1) = 0 \text{ for all } a \ne b. \tag{12}$$

Given this, we assume that the observers actually conduct their measurements according to some partitions. Again, for concreteness, let us assume that those partitions are the ones associated with inputs $x = 0$, $y = 0$. These inputs take on the role of partitions $\mathcal{P}_A$, $\mathcal{P}_B$ in the ontological model picture. The outputs obtained from these measurements are the no-signaling box analogue to the events $\mathcal{P}_A(\omega)$, $\mathcal{P}_B(\omega)$. In order to make the following expressions more concrete, we assume, when $x = 0$, $y = 0$ are input, that the outputs obtained are $a = 0$ and $b = 0$, respectively.

Therefore, given the perfectly correlated events F_A, F_B and numbers $q_A, q_B \in [0, 1]$, we define the sets

$$\alpha_0 = \{a \in \mathcal{A} : p(b = 1|a, x = 0, y = 1) = q_A\}, \quad (13)$$

$$\beta_0 = \{b \in \mathcal{B} : p(a = 1|b, x = 1, y = 0) = q_B\}, \quad (14)$$

and, for all $n \geq 0$,

$$\alpha_{n+1} = \{a \in \alpha_n : p(B_n|a, x = 0, y = 0) = 1\}, \quad (15)$$

$$\beta_{n+1} = \{b \in \beta_n : p(A_n|b, x = 0, y = 0) = 1\}, \quad (16)$$

where

$$A_n = \alpha_n \times \mathcal{B} \times \mathcal{X} \times \mathcal{Y}, \quad (17)$$

$$B_n = \mathcal{A} \times \beta_n \times \mathcal{X} \times \mathcal{Y}. \quad (18)$$

By analogy with the sets in equations (1)–(4), the set α_0 consists of Alice's outcomes such that she assigns probability q_A to F_B upon seeing that outcome, having input $x = 0$, if she assumes that Bob inputs $y = 1$. The set β_1 is the set of Bob's outcomes such that he is certain that Alice assigns probability q_A to F_B upon seeing that outcome, having input $y = 0$, if he assumes that Alice inputs $x = 1$, and so on, and similarly for β_0, α_1, etc.

With this in mind, we can build a chain of mutual certainties, in a manner similar to the classical agreement theorem. If the event $(a = 0, b = 0, x = 0, y = 0)$ is in the sets A_n and B_n, for all $n \in \mathbb{N}$, then, if Alice and Bob both input 0 and get output 0, we have: Alice assigns probability q_A to F_B (assuming that Bob input $y = 1$), Bob is certain that Alice assigns probability q_A to F_B, Alice is certain that Bob is certain that... and so on indefinitely, and vice versa. That is, there is common certainty at $(a = 0, b = 0, x = 0, y = 0)$ that Alice assigns probability q_A to F_B and that Bob assigns probability q_B to F_A.

More formally, there is common certainty about the event that Alice assigns probability q_A to F_B and that Bob assigns probability q_B to F_A if

$$(a = 0, b = 0, x = 0, y = 0) \in A_n \cap B_n \quad \forall n \in \mathbb{N}. \quad (19)$$

There is common certainty of disagreement if, in addition, $q_A \neq q_B$.

Notice the relationship between this definition and the previous one: the ω^* in equation (5), at which the disagreement occurred, fixed the partition elements that Alice and Bob observed. Here, disagreement occurs at the inputs and outputs $(a = 0, b = 0, x = 0, y = 0)$ that the observers obtain.

In complete generality, we can also consider disagreement at arbitrary inputs and outputs (a, b, x, y). For that, one just has to consider the appropriate changes in the preceding paragraphs. The only case that might seem different is that in which Alice and/or Bob's input is the same as that corresponding to the event of interest; that is, $x = 1$ and/or $y = 1$. This is allowed but uninteresting: It is easy to see that the fact that F_A and F_B are perfectly correlated precludes the possibility of common certainty of disagreement in this case. Therefore, for the sake of concreteness, we fix x, y both different from 1 and, in particular, equal to 0.

The following is a rephrasing of the classical agreement theorem:

Theorem 2. Suppose that Alice and Bob share a local no-signaling box with underlying probability distribution p. Let $q_A, q_B \in [0, 1]$, and let

$$p(b = 1|a = 0, x = 0, y = 1) = q_A, \quad (20)$$

$$p(a = 1|b = 0, x = 1, y = 0) = q_B. \quad (21)$$

If q_A and q_B are common certainty between the observers, then $q_A = q_B$.

In Supplementary Note 3, we give a standalone proof of this result. Moreover, using the above correspondence between ontological models and classical no-signaling boxes, one can prove that the notions of common certainty of disagreement in Theorems 1 and 2 are equivalent. We now ask whether this theorem holds in quantum and no-signaling settings.

Observers of quantum systems cannot agree to disagree. Given the mapping exhibited above, as well as the restatement of the agreement theorem for local boxes, it is now natural to ask whether the theorem holds when dropping the locality constraint.

We address this question by exploring it in the broader no-signaling setting. First, we establish that, in general, observers of no-signaling systems can agree to disagree about perfectly correlated events, and we give explicit examples of disagreeing no-signaling distributions. In the particular case of two inputs and two outputs, we characterize the distributions that give rise to common certainty of disagreement. One might think that the fact that observers of no-signaling systems can agree to disagree is a direct consequence of the multitude of uncertainty relations in quantum mechanics, all of which put a limit on the precision with which the values of incompatible observables can be measured and which have even been linked to epistemic inconsistencies in quantum mechanics[18]. Somewhat surprisingly, our next finding shows that this is not the case. We show that disagreeing no-signaling distributions of two inputs and two outputs cannot be quantum—i.e., the agreement theorem holds for observers of quantum systems in this setting. Then, we go beyond this restriction and show that any disagreeing no-signaling distribution with more than two inputs or outputs induces a disagreeing distribution with two inputs and outputs. Since the agreement theorem holds for observers of quantum systems sharing distributions of two inputs and outputs, it does so for more general distributions too. Thus, even if quantum mechanics features uncertainty relations, this does not apply to observers' estimates of perfectly correlated events.

We first present the following theorem in which the no-signaling box has two inputs and two outputs, but we will show in Theorem 5 that the result is fully general. In place of "common certainty of disagreement about the event that Alice assigns probability q_A to $F_B = \{(a, 1, x, 1) : a \in \mathcal{A}, x \in \mathcal{X}\}$ and Bob assigns probability q_B to $F_A = \{(1, b, 1, y) : b \in \mathcal{B}, y \in \mathcal{Y}\}$, at event $(0, 0, 0, 0)$," we simply say "common certainty of disagreement."

Theorem 3. A two-input two-output no-signaling box gives rise to common certainty of disagreement if and only if it takes the form of Table 1.

Table 1 Parametrization of two-input two-output no-signaling boxes with common certainty of disagreement. Here, $r, s, t, u \in [0, 1]$ are such that all the entries of the box are non-negative, $r > 0$, and $s - u \neq r - t$.

$xy \backslash ab$	00	01	10	11
00	r	0	0	$1 - r$
01	$r - s$	s	$-r + t + s$	$1 - t - s$
10	$t - u$	u	$r - t + u$	$1 - r - u$
11	t	0	0	$1 - t$

While some no-signaling distributions can exhibit common certainty of disagreement, we find that probability distributions arising in quantum mechanics do satisfy the agreement theorem. This is surprising: It is well-known that a given measurement of a quantum system (say, that corresponding to the input $x, y = 0$) need not offer any information about the outcome of an incompatible measurement on the same system (say, $x, y = 1$). However, some consistency remains: common certainty of disagreement is impossible, even for incompatible measurements.

Theorem 4. No two-input two-output quantum box can give rise to common certainty of disagreement.

The proof follows by deriving a contradiction from Tsirelson's theorem[19].

We have seen that no two-input two-output quantum box can give rise to common certainty of disagreement. We now lift the restriction on the number of inputs and outputs and show that no quantum box can give rise to common certainty of disagreement.

First, as we now show, the proof for two inputs and outputs does not require common certainty, but only first-order mutual certainty. By observing the definitions of the sets α_n, β_n, one can see that $\alpha_n = \alpha_1$ and $\beta_n = \beta_1$ for all $n \geq 1$. This means that first-order mutual certainty implies common certainty, and, therefore, first-order certainty suffices to characterize the no-signaling box that displays common certainty of disagreement.

As the number of outputs grows, first-order mutual certainty is no longer sufficient. However, since the number of outputs is always finite, there exists an $N \in \mathbb{N}$ such that $\alpha_n = \alpha_N$ and $\beta_n = \beta_N$ for all $n \geq N$. Since $\alpha_{n+1} \subseteq \alpha_n \, \forall \, n$, and similarly for β, the sets α_N, β_N are the smallest sets of outputs for which the disagreement occurs. Because of this, any (a, b, x, y) that belongs to $A_N \cap B_N$ will also belong to $A_n \cap B_n$ for all n; that is, Nth-order mutual certainty implies common certainty. So, for any finite no-signaling box, one needs only Nth-order mutual certainty to characterize it. As the number of outputs grows unboundedly, one needs common certainty to hold[20]. These observations will be relevant to extending Theorem 4 beyond two inputs and outputs.

Theorem 5. No quantum box can give rise to common certainty of disagreement.

To prove the theorem, we show that any no-signaling box with common certainty of disagreement induces a two-input two-output no-signaling box with the same property. Thus, if there existed a quantum system that could generate the bigger box, it could also generate the smaller box. Then, Theorem 3 implies that no quantum box can give rise to common certainty of disagreement.

Observers of quantum systems cannot disagree singularly. Next, we ask if observers of no-signaling quantum systems can disagree in other ways. We define a new notion of disagreement, which we call singular disagreement, by removing the requirement of common certainty and, instead, imposing $q_A = 1, q_B = 0$. We ask whether this new notion holds for observers of classical, quantum, and no-signaling systems. We find the same pattern as before: Singular disagreement does not hold for observers of classical or quantum systems, but can occur in no-signaling settings, where we characterize the distributions that feature it.

Suppose that Alice and Bob share a no-signaling box. As before, Alice assigns probability q_A to the event F_B, and Bob assigns probability q_B to the event F_A. Suppose that this happens at event $(a = 0, b = 0, x = 0, y = 0)$. If $q_A = 1$ and $q_B = 0$, these probabilities differ maximally: Alice is certain that F_B happens, while Bob is certain that F_A does not happen. If, in addition, F_A and F_B are perfectly correlated, then there is singular disagreement at $(0, 0, 0, 0)$. This time, there is no need for chained

Table 2 Parametrization of two-input two-output no-signaling boxes with singular disagreement. Here, $r, s, t, u, \in [0, 1]$ are such that all the entries of the box are non-negative, $s > 0$, and $s + t \neq 0$ and $u + t \neq 1$.

xy\ab	00	01	10	11
00	s	t	$1-s-u-t$	u
01	0	$s+t$	r	$1-s-t-r$
10	$1-u-t$	$u+t+r-1$	0	$1-r$
11	r	0	0	$1-r$

certainties—we just require that Alice's and Bob's assignments differ maximally.

More formally, there is singular disagreement about the probabilities assigned by Alice and Bob to perfectly correlated events $F_B = \{(a, 1, x, 1) : a \in \mathcal{A}, x \in \mathcal{X}\}$ and $F_A = \{(1, b, 1, y) : b \in \mathcal{B}, y \in \mathcal{Y}\}$, respectively, at event $(0, 0, 0, 0)$ if it holds that

$$q_A = 1, \quad q_B = 0. \tag{22}$$

Similarly to the previous section, we refer to the above definition simply as "singular disagreement."

We restrict ourselves first to boxes of two inputs and outputs and show that local boxes cannot exhibit singular disagreement. Then, we characterize the no-signaling boxes that do satisfy singular disagreement and show they cannot be quantum. Finally, we generalize to boxes of any number of inputs and outputs.

Theorem 6. There is no local two-input two-output box that gives rise to singular disagreement.

It turns out that singular disagreement induces a Hardy paradox[21] in the system; therefore it cannot be local.

We now lift the local restriction and characterize the no-signaling boxes in which singular disagreement occurs.

Theorem 7. A two-input two-output no-signaling box gives rise to singular disagreement if and only if it takes the form of Table 2.

The definition of singular disagreement gives rise to the zeros in the second and third rows of Table 2, while the zeros in the bottom row come from the perfectly correlated outputs on inputs $x = y = 1$.

However, singular disagreement cannot arise in quantum systems. This is another way in which quantum mechanics provides some consistency between (possibly incompatible) measurements, just as in the case of common certainty of disagreement.

Theorem 8. No two-input two-output quantum box can give rise to singular disagreement.

By ref. [22], the boxes of Theorem 7 are either local or postquantum. But Theorem 6 implies they cannot be local.

Finally, the above results can be generalized to any finite box:

Theorem 9. No quantum box can give rise to singular disagreement.

The proof is very similar to that of Theorem 5.

Additionally, the PR box[3] satisfies both Theorems 3 and 7, and this makes it an example of both kinds of disagreement.

Discussion

We have defined two notions of disagreement inspired by notions from epistemics. Both notions of disagreement imply immediate tests for new theories—namely, the tables in Theorem 3 and Theorem 7. These tests are very general in the sense that they are based only on the capability (or not) of a theory to realize undesirable correlations between non-communicating parties. Also, both principles have their roots in epistemics, with common

certainty of disagreement closer to Aumann's original idea and singular disagreement permitting a simpler description.

These two notions of disagreement are compatible in that it is possible to find examples displaying both kinds of disagreement at once. Strikingly, a prime such example is the Popescu-Rohrlich box[3], establishing that it is an extremal resource both in the sense of being an extreme point of the polytope of no-signaling distributions and in the sense of inducing the strongest possible disagreement between two parties.

On a speculative note, we suggest that it would be also very interesting to explore the application of the concepts introduced in this paper to practical tasks in which consensus between parties plays a role, such as the coordination of the action of distributed agents or the verification of distributed computations. See ref. [23] for some specific connections along these lines in the classical case.

Further work could be dedicated to constructing physical paradoxes arising from the possibility of agreeing to disagree. The examples in the literature (see refs. [9–11]), while very practical, might, to some communities, be considered less appealing than deeper physical consequences. These could be explored by exploiting the newly built bridges between quantum mechanics and epistemics.

Our results suggest that agreement can be used to design experiments to test the behavior of Nature. In experimental settings, noise is unavoidable. Adding white noise to the boxes in Tables 1 and 2 (both of which lie in quantum voids) would mean that the zeros in the boxes now become small but nonzero parameters. Robustness of quantum voids to this type of noise can be deduced from the closure of the set of quantum 2-input 2-output correlations[24]. This already covers an approximate version of singular disagreement. Another future direction to explore would be defining notions of approximate common certainty of disagreement.

We contend that agreement between observers could be a convenient principle for testing the consistency of new post-quantum theories. Our results yield a clear parameterization of the set of the probability distributions that allow observer disagreement. This set is easy to work with, thanks to its restriction to two observations and two outcomes per observer. If a new theory can be used to generate such a distribution, this might raise a red flag, since this theory violates a reasonable, intuitive and, importantly, testable property that quantum mechanics satisfies.

Therefore, another future direction would be to study disagreement in theories that generalize quantum theory. For instance, one could consider almost quantum correlations[25], which is a set of correlations strictly larger than those achievable by measuring quantum states but that were designed to satisfy all physical principles previously proposed in the literature. Almost quantum correlations are well characterized in terms of no-signaling boxes, so a natural question is whether they permit common certainty of disagreement or singular disagreement. For common certainty of disagreement, a straightforward modification to the proof of Theorem 4 shows that this phenomenon cannot arise under almost quantum correlations. For singular disagreement, the simple characterization of almost quantum correlations in terms of a semidefinite program[26] makes it is possible to search numerically for almost quantum boxes displaying singular disagreement. Using this approach, we have found numerical evidence that singular disagreement, too, cannot arise under almost quantum correlations. Hence, our principles would appear to make it possible to identify additional features that almost quantum correlations share with quantum theory, indicating a new sense in which these generalized correlations are still reasonable.

Furthermore, one of the aims of studying physical principles is to find out whether quantum theory can be deduced from a set of such principles. Our findings open new directions for the exploration of such a programme. However, there are two main challenges in doing so. First, in order to assess this, it would be desirable to compare this principle to existing ones in the literature[3–7]. However, our principle is markedly different in nature—this novelty being precisely the main obstacle. Second, ref. [27] shows that bipartite principles are not enough to capture the set of quantum correlations. This highlights the interest of finding a multipartite extension for our principle. In the classical case, multipartite extensions have already been considered (see, e.g. ref. [28]). Studying similar extensions in the quantum and no-signaling settings is an interesting avenue for future work.

Finally, we compare our results with others in the physics literature that call on Aumann's theorem. References [29,30] examine Aumann's theorem when observers are assumed to use Born's rule to update probabilities. The authors conclude that Aumann's theorem does not hold for this type of observer. Instead, our setting assumes that the observers are macroscopic and merely share a quantum state or a no-signaling box. Our setting is appealing in quantum information for its applications to communication complexity, cryptography, teleportation, and many other scenarios. Reference [31] introduces a different notion of disagreement in a no-signaling context. Here, disagreement concerns pieces of information about some variables, and agreement refers to consistency in the information provided about the variables. Hence, it is unrelated to the epistemic notion of disagreement in Aumann's theorem or in the present work. Disagreement is also a feature of ref. [18]. However, there, one of the observers ignores the fact that a certain measurement has taken place, while another observer takes this into account, and disagreement arises over the outcomes of a different measurement. In our work, the events that take place are the same according to all observers.

Data availability

References

1. Spekkens, R. W. Evidence for the epistemic view of quantum states: a toy theory. *Phys. Rev. A* **75**(March), 032110 (2007).
2. Larsson, J.-Å A contextual extension of Spekkens' toy model. *AIP Conf. Proc.* **1424**, 211–220 (2012).
3. Popescu, S. & Rohrlich, D. Quantum nonlocality as an axiom. *Found. Phys.* **24**(March), 379–385 (1994).
4. Clifton, R., Bub, J. & Halvorson, H. Characterizing quantum theory in terms of information-theoretic constraints. *Found. Phys.* **33**, 1561–1591 (2003).
5. Pawlowski, M. et al. Information causality as a physical principle. *Nature* **461**(October), 1101–1104 (2009).
6. Sun, L.-L., Zhou, X. & Yu, S. No disturbance without uncertainty as a physical principle. Preprint at https://arxiv.org/abs/1906.11807 (2019).
7. Yan, B. Quantum correlations are tightly bound by the exclusivity principle. *Phys. Rev. Lett.* **110**(June), 260406 (2013).
8. Aumann, R. J. Agreeing to disagree. *Ann. Stat.* **4**, 1236–1239 (1976).
9. Sebenius, J. K. & Geanakoplos, J. Don't bet on it: contingent agreements with asymmetric information. *J. Am. Stat. Assoc.* **78**, 424–426 (1983).
10. Milgrom, P. & Stokey, N. Information, trade and common knowledge. *J. Economic Theory* **26**, 17–27 (1982).
11. Aumann, R. J. & Brandenburger, A. Epistemic conditions for nash equilibrium. *Econometrica* **63**, 1161–1180 (1995).
12. Bell, J. On the Einstein Podolsky Rosen paradox. *Phys. Phys. Fiz.* **1**(November), 195–200 (1964).

13. Barrett, J. et al. Nonlocal correlations as an information-theoretic resource. *Phys. Rev. A* **71**(Feb), 022101 (2005).

14. Ferrie, C. Quasi-probability representations of quantum theory with applications to quantum information science. *Rep. Prog. Phys.* **74**, 116001 (2011).

15. Aaronson, S. The complexity of agreement. In *Proceedings of the Thirty-Seventh Annual ACM Symposium on Theory of Computing*, STOC '05, pages 634–643, New York, NY, USA, 2005. Association for Computing Machinery.

16. Abramsky, S. & Brandenburger, A. The sheaf-theoretic structure of non-locality and contextuality. *N. J. Phys.* **13**(November), 113036 (2011).

17. Abramsky, S. and Brandenburger, A. In *Horizons of the Mind. A Tribute to Prakash Panangaden* (eds van Breugel, F., Kashefi, E., Palamidessi, C. & Rutten, J.) (Springer International Publishing, 2014).

18. Frauchiger, D. & Renner, R. Quantum theory cannot consistently describe the use of itself. *Nat. Commun.* **9**, 3711 (2018).

19. Cirel'son, B. S. Quantum generalizations of Bell's inequality. *Lett. Math. Phys.* **4**, 93–100 (1980).

20. Geanakoplos, J. D. & Polemarchakis, H. M. We can't disagree forever. *J. Econ. Theory* **28**, 192–200 (1982).

21. Hardy, L. Quantum mechanics, local realistic theories, and Lorentz-invariant realistic theories. *Phys. Rev. Lett.* **68**(May), 2981–2984 (1992).

22. Rai, A., Duarte, C., Brito, S. & Chaves, R. Geometry of the quantum set on no-signaling faces. *Phys. Rev. A* **99**(March), 032106 (2019).

23. Fagin, R., Halpern, J., Moses, Y. & Vardi, M. *Reasoning About Knowledge* (MIT Press, 2003).

24. Goh, K. T. et al. Geometry of the set of quantum correlations. *Phys. Rev. A* **97**(Feb), 022104 (2018).

25. Navascués, M., Guryanova, Y., Hoban, M. J. & Acín, A. Almost quantum correlations. *Nat. Commun.* **6**, (2015).

26. Navascués, M., Pironio, S. & Acín, A. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *N. J. Phys.* **10**(jul), 073013 (2008).

27. Gallego, R., Würflinger, L. E., Acín, A. & Navascués, M. Quantum correlations require multipartite information principles. *Phys. Rev. Lett.* **107**(Nov), 210403 (2011).

28. Parikh, R. & Krasucki, P. Communication, consensus, and knowledge. *J. Econ. Theory* **52**, 178–189 (1990).

29. Khrennikov, A. Quantum version of Aumann's approach to common knowledge: Sufficient conditions of impossibility to agree on disagree. *J. Math. Econ.* **60**(October), 89–104 (2015).

30. Khrennikov, A. & Basieva, I. Possibility to agree on disagree from quantum information and decision making. *J. Math. Psychol.* **62-63**(October), 1–15 (2014).

31. Abramsky, S. & Carù, G. Non-locality, contextuality and valuation algebras: a general theory of disagreement. *Philos. Trans. R. Soc. A* **377**(November), 20190036 (2019).

Author contributions

P.C.-T. and G.S. contributed equally to this work. P.C.-T., G.S., A.M.K., A.B., and P.L.M. contributed to designing the ideas, performing the calculations, analyzing the results, and writing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-27134-6.

Correspondence and requests for materials should be addressed to Giannicola Scarpa.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.