Measurement sharpness cuts nonlocality and contextuality in every physical theory

Giulio Chiribella and Xiao Yuan

Center for Quantum Information, Institute for Interdisciplinary
Information Sciences, Tsinghua University, Beijing, 100084, China

(Dated: April 21, 2014)

Gathering data through measurements is at the basis of every experimental science. Ideally, measurements should be repeatable and, when extracting only coarse-grained data, they should allow the experimenter to retrieve the finer details at a later time. However, in practice most measurements appear to be noisy. Here we postulate that, despite the imperfections observed in real life experiments, there exists a fundamental level where all measurements are ideal. Combined with the requirement that ideal measurements remain so when coarse-grained or applied in parallel on spacelike separated systems, our postulate places a powerful constraint on the amount of nonlocality and contextuality that can be found in an arbitrary physical theory, bringing down the violation of Bell and Kocher-Specker inequalities near to its quantum value. In addition, it provides a new compelling motivation for the principles of Local Orthogonality and Consistent Exclusivity, recently proposed for the characterization of the quantum set of probability distributions.

Nonlocality [1–3] and contextuality [4, 5] are among the most striking features of quantum mechanics, in radical conflict with the worldview of classical physics. Still, quantum mechanics is neither the most nonlocal theory one can imagine, nor the most contextual. For nonlocality, this observation dates back to the seminal work of Popescu and Rohrlich [6], who showed that relativistic no-signalling is compatible with correlations that are much stronger than those allowed by quantum theory. Their work stimulated the question whether other fundamental principles, yet to be discovered, characterize the peculiar set of correlations observed in the quantum world. Up to now, several candidates that partly retrieve the set of quantum correlations have been proposed, including Non-Trivial Communication Complexity [7, 8], No-Advantage in Nonlocal Computation [9], Information Causality [10], Macroscopic Locality [11], and, lately, Local Orthogonality (LO) [12]. The observation that quantum theory is not maximally contextual made an early appearance in Kochen's and Specker's work [4, 13], but it was not until recently that it caught broad attention in the community [14–16], initiating a search for principles that characterize the quantum set of contextual probability distributions. On this front, the only principle put forward so far is Consistent Exclusivity (CE) [15, 17–19].

Despite many successes, a complete characterization of the quantum set is still extremely challenging [20]. What makes the problem hard is the fact that—intendedly— the principles considered so far dealt only with input-output probability distributions, without making any hypothesis on how these distributions are generated. On the other hand, a physical theory does not provide only probability distributions, but also specifies rules on how to combine physical systems together, how to measure them, and how to evolve their state in time [21]. Considering that fundamental quantum features like no-cloning and universal computation cannot be expressed just in terms of input-output distributions, it is natural to wonder whether also quantum nonlocality and contextuality could be better understood in a broader framework of general probabilistic theories (GPTs) [22–25]. Further motivation to extend the framework comes from the latest principles in the nonlocality and contextuality camps: LO and CE. Both principles refer to a notion of orthogonal events and impose that the sum of the probabilities of a set of mutually orthogonal events shall not exceed one. This is a powerful requirement, which in the case of LO is even capable to rule out non-quantum correlations that are compatible with every bipartite principle [26]. But why should Nature obey such a requirement? And what does this requirement tell us about the fundamental laws that govern physical processes?

Here we tackle the problem of understanding quantum nonlocality and contextuality from a new angle, which focuses on the fundamental structure of measurements in arbitrary physical theories. We introduce a class of ideal measurements, called sharp, that are repeatable and cause the minimal amount of disturbance on future observations. We postulate that all measurements are sharp at the fundamental level and we explain the apparent unsharpness of realistic experiments as due to the interaction with the environment surrounding the measured system. Assuming that sharp measurements remain sharp under elementary operations, such as joining two outcomes together and applying two measurements in parallel, we show that the fundamental sharpness of measurements implies the validity of CE and LO, thus providing a strong constraint on the set of probability distributions and bringing down the violation of Bell and Kochen-Specker inequalities near to their quantum value. Our result demonstrates that principles formulated in the broader framework of GPTs can offer an extra power in the characterization of the quantum set and identifies the fundamental sharpness of measurements as a candidate principle for future axiomatizations of quantum theory.

I. RESULTS

Framework. In a general physical theory, a measure-

ment is described by a collection of events, each event labelled by an outcome. We first consider demolition measurements, which adsorb the measured system. In this case, the measurement events are called *effects* and the measurement is a collection of effects $\{m_x\}_{x \in \mathsf{X}}$. For a system prepared in the state ρ, the probability of the outcome x is denoted by $p_x = (m_x|\rho)$. In quantum theory this is a notation for the Born rule $p_x = \mathrm{Tr}[m_x\rho]$, where ρ is a density matrix and m_x is a measurement operator. In general theories, $(m_x|\rho)$ does not denote a trace of matrices and in fact the actual recipe for computing the probability $(m_x|\rho)$ is irrelevant here. We will often use the notation $(m_x|$ and $|\rho)$ for effects and states, respectively. It is understood that two different states give different probabilities for at least one effect, and two different effects take place with different probabilities on at least one state.

When two measurements $\{m_x\}$ and $\{n_y\}$ are performed in parallel on two systems A and B, we denote by $m_x \otimes n_y$ the measurement event labelled by the pair of outcomes x, y. Similarly, when two states of systems A and B, say α and β, are prepared independently, we denote by $\alpha \otimes \beta$ the corresponding state of the composite system AB. In quantum theory, this is the ordinary tensor product of operators, but this may not be the case in a general theory and, again, the actual recipe for computing $\alpha \otimes \beta$ is irrelevant here. What is relevant, instead, is that the notation is consistent with the operational notion of performing independent operations on different systems: If two systems are independently prepared in states α and β and undergo to independent measurements $\{m_x\}$ and $\{n_y\}$, we impose that the probability has the product form $p_{xy} = (m_x|\alpha)(n_y|\beta)$.

The most basic operation one can perform on a measurement is to join some outcomes together, thus obtaining a new, less informative measurement. This operation, known as *coarse-graining*, is achieved by dividing the outcomes of the original measurement $\{m_x\}_{x \in \mathsf{X}}$ into disjoint groups $\{\mathsf{X}_z\}_{z \in \mathsf{Z}}$, and by identifying outcomes that belong to the same group. The result of this procedure is a new measurement $\{m'_z\}_{z \in \mathsf{Z}}$ satisfying the relation $(m'_z|\rho) = \sum_{x \in \mathsf{X}_z}(m_x|\rho)$ for every every z and for every possible state ρ. For brevity, we write $m'_z = \sum_{x \in \mathsf{X}_z} m_x$.

Coarse-graining allows one to express the principle of *causality*, which states that the settings of future measurements do not influence the outcome probabilities of present experiments [24]. Causality is equivalent to the requirement that for every system A there exists an effect u_A, called the *unit*, such that

$$\sum_{x \in \mathsf{X}} m_x = u_A \tag{1}$$

for every measurement $\{m_x\}_{x \in \mathsf{X}}$ on A. In quantum theory, u_A is the identity operator on the Hilbert space of the system and Eq. (1) expresses the fact that quantum measurements are resolutions of the identity. When there is no ambiguity, we drop the subscript from u_A.

Causality has major consequences. First of all, it implies that the probability distributions generated by local measurements satisfy the no-signalling principle [24]. Moreover, it allows to perform adaptive operations: for example, if $\{m_x\}_{x \in \mathsf{X}}$ is a measurement on system A and $\{n_y^{(x)}\}_{y \in \mathsf{Y}}$ is a measurement on system B for every value of x, then causality guarantees that it is possible to choose the measurement on B depending on the outcome on system A, i. e. that $\{m_x \otimes n_y^{(x)}\}_{x \in \mathsf{X}, y \in \mathsf{Y}}$ is a legitimate measurement. Finally, causality allows one to describe non-demolition measurements. For a non-demolition measurement $\{\mathcal{M}_x\}_{x \in \mathsf{X}}$, the measurement events are *transformations*, which turn the initial state of the system, say ρ, into a new unnormalized state $\mathcal{M}_x|\rho)$. For a system prepared in the state ρ, the probability of the outcome x is $p_x = (u|\mathcal{M}_x|\rho)$ and, conditionally on outcome x, the post-measurement state is $\mathcal{M}_x|\rho)/(u|\mathcal{M}_x|\rho)$. We will often refer to the non-demolition measurements as *instruments*, in analogy with the usage in quantum theory [27, 28]. Note that, thanks to causality, every instrument $\{\mathcal{M}_x\}$ is associated to a unique demolition measurement $\{m_x\}$ via the relation

$$(m_x| = (u|\mathcal{M}_x \qquad \forall x \in \mathsf{X}. \tag{2}$$

By definition, $\{m_x\}$ describes the statistics of the instrument: for every state ρ and for every outcome x, one has $p_x = (u|\mathcal{M}_x|\rho) \equiv (m_x|\rho)$.

Sharp measurements in arbitrary theories. In textbook quantum mechanics, physical quantities are associated to self-adjoint operators, called observables [29]. The values of a quantity are the eigenvalues of the corresponding operator and the probability that a measurement outputs the value x is given by the Born rule $p_x = \mathrm{Tr}[P_x\rho]$, where P_x is the projector on the eigenspace for the eigenvalue x and ρ is the density matrix of the system before the measurement. If the measurement gives the outcome x, then the state after the measurement is $\rho'_x = P_x\rho P_x/\mathrm{Tr}[P_x\rho]$, according to the projection postulate. These canonical measurements, where all the measurement operators are orthogonal projectors, are called sharp [30]. While it is clear that sharp measurements play a key role in quantum theory it is by far less clear how to define them in an arbitrary GPT. Here we propose a simple definition based on the notions of repeatability and minimal disturbance.

Let us start from repeatability. An instrument $\{\mathcal{M}_x\}$ is *repeatable* if it gives the same outcome when performed two consecutive times, namely

$$(m_x|\mathcal{M}_x = (m_x| \qquad \forall x \in \mathsf{X}, \tag{3}$$

where $\{m_x\}$ is the measurement of Eq. (2). Repeatability poses a fairly weak requirement on $\{m_x\}$: every measurement that discriminates perfectly among a set of states $\{\rho_x\}$ can be realized by a repeatable instrument, which consists in measuring $\{m_x\}$ and, if the outcome is x, re-preparing the system in state ρ_x.

The second ingredient entering in our definition of sharp measurements is minimal disturbance. We say that the instrument $\{\mathcal{M}_x\}_{x\in\mathsf{X}}$ does not disturb the measurement $\mathbf{n} = \{n_y\}_{y\in\mathsf{Y}}$ if the former does not affect the statistics of the latter, namely

$$(n_y|\mathcal{M} = (n_y| \qquad \forall y \in \mathsf{Y}\,, \qquad (4)$$

where $(n_y|\mathcal{M} := \sum_{x\in\mathsf{X}}(n_y|\mathcal{M}_x$. Then, we ask which instruments disturb the smallest possible set of measurements. Clearly, if $\{\mathcal{M}_x\}$ does not disturb $\mathbf{n}$, then $\mathbf{n}$ must be compatible with the measurement $\mathbf{m} = \{(u|\mathcal{M}_x\}$, in the sense that $\mathbf{m}$ and $\mathbf{n}$ can be measured jointly. Indeed, by measuring $\mathbf{n}$ after $\{\mathcal{M}_x\}$ one obtains the probability distribution $p_{xy} = (n_y|\mathcal{M}_x|\rho)$, whose marginals on x and y are equal to the probability distributions of $\mathbf{m}$ and $\mathbf{n}$, respectively. Read in the contrapositive, this means that if $\mathbf{m}$ and $\mathbf{n}$ are incompatible, the instrument $\{\mathcal{M}_x\}$ must disturb $\mathbf{n}$. This leads us to the following definition: an instrument $\{\mathcal{M}_x\}$ has *minimal disturbance* if it disturbs only the measurements that are incompatible with $\mathbf{m} = \{(u|\mathcal{M}_x\}$.

We define an instrument to be sharp if it is both repeatable and with minimal disturbance. We say that a measurement is sharp if it describes the statistics of a sharp instrument and we call an effect sharp if it belongs to a sharp measurement. In quantum theory, our definition coincides with the usual one: one can prove that the only sharp instruments are the Lüders instruments [31], of the form $\mathcal{M}_x(\rho) = P_x\rho P_x$ where $\{P_x\}$ is a collection of orthogonal projectors. Hence, the sharp measurements are projective measurements. In addition, we can prove that when a sharp measurement extracts a coarse-grained information, the experimenter can still retrieve the finer details at a later time. In fact, this is a necessary and sufficient condition for a measurement to be sharp, as proven in the Methods section.

Fundamental sharpness of measurements. Sharp measurements are an ideal standard—they are the measurements that generate outcomes in a repeatable way, while at the same time causing the least disturbance on future observations. Unfortunately though, most measurements in real life appear to be noisy and not repeatable. Hence the natural question: Is noise fundamental? Or rather it is contingent to the fact that the experimenter has incomplete control on the conditions of the experiment? Here we state that noise is not fundamental and only arises from the fact that the realistic measurements do not extract information only from the system, but also from the surrounding environment:

Principle 1 (Fundamental Sharpness of Measurements)**.** *Every measurement arises from a sharp measurement performed jointly on the system and on the environment.*

Precisely, we require that for every measurement $\mathbf{m} = \{m_x\}_{x\in\mathsf{X}}$ there exists an environment E, a state σ of E, and a sharp measurement $\mathbf{M} = \{M_x\}_{x\in\mathsf{X}}$ on the composite system SE such that, for every state ρ of system S, one has $(m_x|\rho) = (M_x|\,\rho\otimes\sigma)$ for every outcome $x \in \mathsf{X}$. In quantum theory, this is the content of the celebrated Naimark's theorem [32, 33]. This is a deep property, hinting at the idea there exists a fundamental level where all measurements are ideal.

Let us push the idea further. If measurements are sharp at the fundamental level, it is natural to assume that the set of sharp measurements is closed under the basic operation of coarse-graining, which transforms an initial measurement $\mathbf{m}$ into a new, less informative measurement $\mathbf{m}'$. Indeed, since $\mathbf{m}'$ provides less information than $\mathbf{m}$, one expects that $\mathbf{m}'$ should not be less repeatable, nor create more disturbance, than $\mathbf{m}$. This intuition leads to the following requirement:

Principle 2 (Less Information, More Sharpness)**.** *If a measurement is less informative than a sharp measurement, then it is sharp.*

Suppose now that two experimenters, Alice and Bob, perform two sharp measurements on two systems A and B in their laboratories. Again, if measurements are sharp at the fundamental level, one expects the result of Alice's and Bob's measurements to be a sharp measurement on the composite system AB. If this were not the case, it would mean that at the fundamental level some measurements require nonlocal interactions, even though at the operational level the they appear to be implemented locally by Alice and Bob. We then postulate the following

Principle 3 (Locality of Sharp Measurements)**.** *If two sharp measurements are applied in parallel on systems A and B, then the result is a sharp measurement on the composite system AB.*

Principles 1-3 lay down the fundamental structure of sharp measurements, summarized in Fig. 1. They are satisfied by classical theory and by quantum theory, both on complex and real Hilbert spaces. In the following we will show that the fundamental structure of sharp measurements has an enormous impact on the amount of nonlocality and contextuality that can be found in a physical theory.

Derivation of CE. At present, CE is the only principle known to constrain the amount of contextuality of a generic theory. Operationally, the principle can be formulated as follows: Consider a collection of sharp measurements $\{\mathbf{m}^{(x)}, x \in \mathsf{X}\}$, each measurement having outcomes in a set Y_x. Suppose that the possible events have been labelled so that two effects corresponding to the same outcome coincide, i. e. $m_y^{(x)} \equiv m_y$, independently of x. Letting $\mathsf{Y} = \cup_x\mathsf{Y}_x$ be the set of all outcomes, one calls two distinct outcomes $y, y' \in \mathsf{Y}$ *exclusive* if there exists a measurement setting x such that both y and y' belong to Y_x. We say that a theory satisfies CE if for every set of mutually exclusive outcomes E and for every state ρ the probabilities $p_y = (m_y|\rho)$ obey the bound $\sum_{y\in\mathsf{E}} p_y \leq 1$.
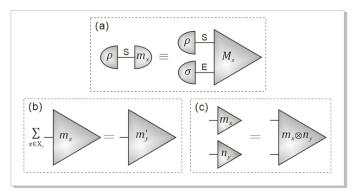
FIG. 1. **The fundamental structure of sharp measurements.** (**a**) Every non-sharp measurement $\{m_x\}_{x \in \mathsf{X}}$ (round diagram on the l.h.s.) performed on system S in state ρ is equivalent to a sharp measurement $\{M_x\}_{x \in \mathsf{X}}$ (triangular diagram on the r.h.s.) performed jointly on the system and on an environment E in state σ. (**b**) Coarse-graining a sharp measurement $\{m_x\}_{x \in \mathsf{X}}$ yields a new sharp measurement $\{m'_y\}_{y \in \mathsf{Y}}$. (**c**) When two sharp measurements $\{m_x\}$ and $\{m_y\}$ are performed in parallel, they yield a new sharp measurement $\{m_x \otimes n_y\}$.

Our first key result is the derivation of CE. In fact, we prove a stronger result: We define two sharp effects m and m' to be *orthogonal* if they belong to the same (not necessarily sharp) measurement and we prove that mutually orthogonal effects can be combined into a single sharp measurement (see Methods). Clearly, since mutually exclusive outcomes correspond to mutually orthogonal effects, the existence of a joint measurement containing the effects $\{m_y\}_{y \in \mathsf{E}}$ implies the bound $\sum_{y \in \mathsf{E}} (m_y|\rho) \leq 1$. Our result implies that in a theory where measurements are fundamentally sharp the violation of Kochen-Specker inequalities is upper bounded by the value set by CE [16]. What is remarkable here is that a single requirement on measurements influences directly the strength of contextuality in an arbitrary physical theory. This situation contrasts with that of the known axiomatizations of quantum theory [22, 34–38], where the quantum bounds on contextuality are retrieved only indirectly through the derivation of the Hilbert space framework.

Our principles do not imply only CE, but also the whole hierarchy of extensions of this principle defined in Ref. [18]. The L-th level of the hierarchy can be defined by considering independent measurements on L copies of the state ρ. Denoting by $\mathbf{y} = (y_1, \dots, y_L)$ the string of all outcomes, one says that two strings $\mathbf{y}$ and $\mathbf{y}'$ are exclusive if there exists some i such that y_i and y'_i are exclusive. A physical theory satisfies the L-th level of the hierarchy if the probabilities $p_L(\mathbf{y}) = \prod_{i=1}^{L} (m_{y_i}|\rho)$ obey the bound $\sum_{\mathbf{y} \in \mathsf{E}} p_L(\mathbf{y}) \leq 1$ for every set E of mutually exclusive strings. In the Methods section we show that our principles on sharp measurements imply that this bound is satisfied for every possible L. Again, this fact has major

consequences on the amount of contextuality that can be found in a theory satisfying our principles. For example, choosing $L = 2$ and invoking a result of Ref. [15] we have that the structure of sharp measurements implies that the maximum violation of KCBS inequality [39] is exactly equal to the quantum value.

Derivation of LO. In the nonlocality camp, LO occupies a special position, being up to now the only known principle that rules out non-quantum correlations that are not detected by any bipartite principle [26]. LO refers to a scenario where N parties perform local measurements on N systems, initially prepared in some joint state. The i-th party can choose among different measurement settings in a set X_i and her measurements give outcomes in another set Y_i. Let $\mathbf{x} = (x_1, \dots, x_N)$ be the string of all settings, $\mathbf{y} = (y_1, \dots, y_N)$ be the string of all outcomes, and $\mathbf{e}$ be the pair $\mathbf{e} = (\mathbf{x}, \mathbf{y})$. In this context, the pair $\mathbf{e} = (\mathbf{x}, \mathbf{y})$ is called an event and two events are called *locally orthogonal* iff there exists a party i such that $x_i = x'_i$ and $y_i \neq y'_i$. Setting $p(\mathbf{e})$ to be the conditional probability distribution $p(\mathbf{y}|\mathbf{x})$, one says that theory satisfies local orthogonality if all the probability distributions generated by local measurements obey the bound $\sum_{\mathbf{e} \in \mathsf{O}} p(\mathbf{e}) \leq 1$ for every set O of pairwise locally orthogonal events.

To derive LO, we specify how the probability distribution $p(\mathbf{y}|\mathbf{x})$ is generated: In the most general scenario, the N parties share a state ρ and that, for setting x_i, party i performs a measurement $\mathbf{m}^{(i,x_i)}$. Denoting the product effects as $P_{\mathbf{y}}^{(\mathbf{x})} := \bigotimes_{i=1}^{N} m_{y_i}^{(i,x_i)}$, the probability distribution of the outcomes is given by $p(\mathbf{e}) := \left(P_{\mathbf{y}}^{(\mathbf{x})}|\rho\right)$. The proof that LO follows from the principles, provided in Methods, consists of three steps: First, thanks to the Fundamental Sharpness of Measurements, the problem is reduced to proving that LO holds for probability distributions generated by sharp local measurements. Then, we observe that, in the case of sharp measurements, locally orthogonal events correspond to orthogonal effects. Finally, we use the fact that mutually orthogonal effects can coexist in a single measurement. As a corollary, we obtain the bound $\sum_{\mathbf{e} \in \mathsf{O}} p(\mathbf{e}) \leq 1$, establishing the validity of LO for all the probability distributions generated by measurements in our theory.

Like in the case of CE, our principles imply the whole hierarchy of extensions of LO introduced in [12]. The hierarchy is defined as follows: the probabilities $p(\mathbf{y}|\mathbf{x})$ satisfy the L-th level of the hierarchy if their product $p(\mathbf{y}_1|\mathbf{x}_1)\cdots p(\mathbf{y}_L|\mathbf{x}_L)$ satisfies LO. Now, we can think of the product as being generated by measurements on N copies of the state ρ. In this way, we reduce the problem of proving the L-th level of the hierarchy to the problem of proving LO for measurements performed on the state $\rho^{\otimes L}$. But we already proved the validity of LO for arbitrary measurements and arbitrary states. In conclusion, the structure of sharp measurements implies that LO is satisfied at every possible level.

A striking consequence of this argument is that the fun-

damental sharpness of measurements rules out the ultra-strong nonlocality exhibited by PR box correlations, as the latter violate the LO hierarchy [12]. In other words, in the world of PR boxes some measurements must be fundamentally noisy.

Sharp Bell inequalities. The request that measurements are ideal at the fundamental level exerts a censorship on the amount of nonlocality that can be detected by experiments. To illustrate this fact, we show a number of Bell inequalities where the sharpness of measurements prevents every violation. We call such inequalities sharp.

Consider a game played by N non-communicating parties and a referee, who sends to party i an input x_i and receives back an output y_i. The referee chooses the input string $\mathbf{x}$ at random with probability $q(\mathbf{x})$ and assigns a payoff $\omega(\mathbf{x}, \mathbf{y})$ to the players, assumed without loss of generality to be nonnegative for every $(\mathbf{x}, \mathbf{y})$. The expected payoff obtained by the players is given by $\omega = \sum_{\mathbf{x}, \mathbf{y}} q(\mathbf{x})\omega(\mathbf{x}, \mathbf{y})p(\mathbf{y}|\mathbf{x})$, where $p(\mathbf{y}|\mathbf{x})$ is the probability distribution describing their strategy. For a given game, the maximum payoff that can be achieved by classical strategies—call it ω_c—defines a Bell inequality, $\omega \leq \omega_c$. The game can be associated with a graph G, here called the *winning graph*, by choosing as vertices the events $(\mathbf{x}, \mathbf{y})$ such that $q(\mathbf{x})\omega(\mathbf{x}, \mathbf{y}) \neq 0$ and placing an edge between two events $\mathbf{e}$ and $\mathbf{e}'$ if they are not locally orthogonal, as illustrated in Figure 2. In this picture, the maximum payoff achieved by classical strategies is

$$\omega_c = \max_{\mathsf{C} \subseteq \mathsf{G}} \sum_{(\mathbf{x}, \mathbf{y}) \in \mathsf{C}} q(\mathbf{x})\, \omega(\mathbf{x}, \mathbf{y}), \qquad (5)$$

where C is a clique, i. e. a subset of G with the property that every two vertices in C are connected [12].

A first class of games leading to sharp Bell inequalities is the class of games with a graph G that is the disjoint union of mutually disconnected cliques C_k, $k \in \{1, \dots, K\}$. This class contains the game Guess Your Neighbor's Input [40] and the maximally difficult Distributed Guessing Problems of Ref. [12]. In addition, it contains other games such as, for even N, the "Guess the Parity" game where each player is asked to guess the parity of the input string $\mathbf{x}$. For all these games, LO implies that the classical payoff is an upper bound. Indeed, picking for every k the event $\mathbf{e}_k \in \mathsf{C}_k$ that has maximum probability, the payoff can be bounded as

$$\omega = \sum_{k=1}^{K} \sum_{(\mathbf{x}, \mathbf{y}) \in \mathsf{C}_k} q(\mathbf{x})\, \omega(\mathbf{x}, \mathbf{y})\, p(\mathbf{y}|\mathbf{x})$$

$$\leq \sum_{k=1}^{K} p(\mathbf{e}_k) \left[\sum_{(\mathbf{x}, \mathbf{y}) \in \mathsf{C}_k} q(\mathbf{x})\, \omega(\mathbf{x}, \mathbf{y}) \right]$$

and since the events $\{\mathbf{e}_k\}_{k=1}^{K}$ are locally orthogonal by construction, one has $\sum_k p(\mathbf{e}_k) \leq 1$ and, therefore, $\omega \leq \max_k q(\mathsf{C}_k) \equiv \omega_c$. In conclusion, every game where the wining graph is a disjoint union of disconnected cliques
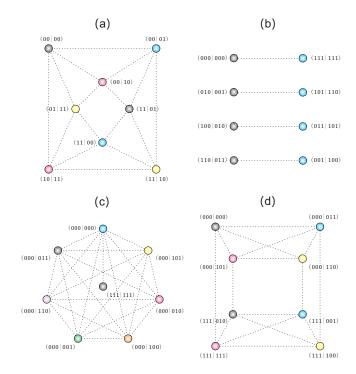

FIG. 2. **Winning graphs.** The vertices are coloured so that two adjacent vertices have distinct colours, using the minimum number of colours. (**a**) Winning graph for the CHSH game [3]. The player win $+1$ if $y_1 \oplus y_2 = x_1 x_2$ and 0 otherwise. The graph is not perfect, because the largest clique in the graph has 3 vertices while the number of colours in the graph is 4. Here classical strategies are not optimal among the strategies that satisfy LO. (**b**) Winning graph for Guess Your Neighbor's Input [40] in the case of $N = 4$ parties. The players win $+1$ if $y_i = x_{i+1}$ for every i. The graph is a disjoint union of disconnected cliques and therefore classical strategies are optimal among all strategies satisfying LO. (**c**) Winning graph for the game "Guess the Product" in the case of $N = 3$ parties. The players win $+1$ if $y_i = x_1 x_2 x_3$ for every i and 0 otherwise. The graph is perfect and therefore the classical strategy is optimal [18]. (**d**) Winning graph for the game "Guess the Parity" in the case of $N = 3$ parties. The players win $+1$ if $y_i = x_1 \oplus x_2 \oplus x_3$ for every i and 0 otherwise. Also in this case, the graph is perfect and the classical strategy is optimal.

defines a Bell inequality, $\omega \leq \omega_c$ that cannot be violated by any theory satisfying LO, and, in particular, by any theory where measurements are fundamentally repeatable and with minimal disturbance. Using a result of Ref. [18], the proof that LO cuts the payoff down to its classical value can be extended to a larger class of games, defined by the property that the winning graph G is a perfect graph [41] (recall that a graph is perfect if, for every subset S of its vertices, the number of colours needed to label adjacent vertices in S with different colours is equal to the number of vertices in the largest clique $\mathsf{C} \subseteq \mathsf{S}$). Examples of such games are the "Guess the Parity" game and the "Guess the Product" game where the players are win if they guess the product of their inputs.

II. DISCUSSION

Our results are derived in a minimal framework, which avoids some assumptions commonly made in GPTs. In particular, our arguments do not invoke local tomography [22, 42, 43], but only the requirement that sharp measurements are local. This requirement is strictly weaker: for example, it is satisfied by quantum theory on real Hilbert spaces, where local tomography fails. The validity of our results does not even require that the states of a given physical system form a convex set. Thanks to this feature, the results apply also to non-convex theories, like Spekkens' toy theory [44]. Interestingly, probabilities themselves do not play a crucial role in our arguments and it is quite straightforward to extend the results to theories that only specify which outcomes are possible, impossible or certain, without specifying their probabilities, such as Schumacher's and Westmoreland's toy theory [45].

Since sharp measurements play a central role in quantum mechanics, it is not surprising that they have been the object of extensive investigation since the early days [31, 46, 47]. Later, Holevo proposed a purely statistical definition of sharp measurement, which does not refer to post-measurement states [48]. Although in the quantum case Holevo's definition reduces to that of projective measurement, in general theories it is inequivalent to ours, and it is not clear how one could use it to derive features like LO and CE. Different notions of ideal measurements were put forward by Piron [49, 50] and Beltrametti-Cassinelli [51] in the framework of quantum logic. In general, they differ form our definition in the way the condition of minimal disturbance is defined. Most recently, measurement disturbance came back to play a key role in the search for basic physical principles, as shown e. g. by the No Disturbance Without Information principle of Ref. [52] and by the No Disturbance principle of Ref. [53].

Our work joined the insights from two different approaches to the foundations of quantum mechanics: the characterization of quantum correlations [6–12] and the study of general probabilistic theories [22, 34–38]. Although these two approaches have developed on separate tracks so far, they share the same fundamental goal: understanding which picture of Nature lies behind the laws of quantum mechanics and guiding our intuition towards the formulation of new physical theories and new information-processing protocols. Our results demonstrate that the interaction between the two approaches can be highly beneficial for both. Here LO and CE stimulated the search for new principles in the GPT framework, leading to a compelling picture of nature where measurements are repeatable and cause minimal disturbance at the fundamental level. It is an open question whether other information-theoretic features of quantum correlations, such as Non-Trivial Communication Complexity, No-Advantage in Nonlocal Computation, and Information Causality can be derived in this picture. Fur-

thermore, the idea that a noisy physical process can be reduced to an ideal process at the fundamental level reminds immediately of another quantum feature: Purification [24]. Operationally, Purification is the property that every mixed state can be generated from a pure state of a composite system by discarding one component. This principle implies directly entanglement and is at the core of the reconstruction of quantum theory of Ref. [34]. Our result suggests the possibility that purification and sharpness could be sufficient to single out quantum theory among all possible GPTs. In terms of quantum correlations, this would lead to the tantalizingly simple picture "Purification brings nonlocality in, sharpness cuts it down". Going even further, it is intriguing to envisage that purification and sharpness could be viewed as two sides of the same medal by imposing that physical theories must satisfy a requirement of time symmetry, similarly to what was done in quantum theory by Aharonov, Bergmann and Lebowitz [54, 55].

III. METHODS

Characterization of sharp instruments. The starting point of our results is the observation that an instrument $\{\mathcal{M}_x\}$ is sharp if and only if

$$(r_{xy}|\mathcal{M}_x = (r_{xy}| \qquad \forall x \in \mathsf{X}, \forall y \in \mathsf{Y} \qquad (6)$$

for every measurement $\mathbf{r} = \{r_{xy}\}_{(x,y)\in\mathsf{X}\times\mathsf{Y}}$ that refines $\{m_x\}$, i. e. $\sum_{y\in\mathsf{Y}} r_{xy} = m_x$ for every x. Let us see why Eq. (6) is equivalent to sharpness. First, suppose that Eq. (6) holds. Clearly, this implies that $\mathbf{m}$ is repeatable, as one can see by summing over y. Moreover, Eq. (6) implies that $\mathbf{m}$ is a minimal disturbance measurement. Indeed, take a generic measurement $\mathbf{n}$ that is compatible with $\mathbf{m}$. By definition, this means that there exists a joint measurement $\mathbf{r} = \{r_{xy}\}$ such that $\sum_x r_{xy} = n_y$ for every y and $\sum_y r_{xy} = m_x$ for every x. We then obtain

$$(n_y|\mathcal{M} = \sum_x (r_{xy}|\mathcal{M}_x + (s_y| \qquad (s_y| := \sum_{x,x':x\neq x'} (j_{xy}|\mathcal{M}_{x'}$$

$$= \sum_x (r_{xy}| + (s_y|$$

$$= (n_y| + (s_y|, \qquad (7)$$

having used Eq. (6) in the second equality. Summing over y and using the normalization of the measurement $\mathbf{n}$ we obtain the condition $\sum_y (s_y| = 0$, or, equivalently, $\sum_y (s_y|\rho) = 0$ for every ρ. Since probabilities are nonnegative, this implies that each term in the sum vanishes, leading to the relation $s_y = 0$ for every y. Inserting this relation back in Eq. (7) we conclude that $(n_y|\mathcal{M} = (n_y|$, that is, the instrument does not disturb $\mathbf{n}$. Hence, Eq. (6) implies that $\{\mathcal{M}_x\}$ is a sharp instrument. Conversely, if $\{\mathcal{M}_x\}$ is a sharp instrument then Eq. (6) must be satisfied. By definition, one has

$$(m_x| = (e|\mathcal{M}_x = \sum_{x'} (m_{x'}|\mathcal{M}_x$$

and using the repeatability condition $(m_x| = (m_x|\mathcal{M}_x$ one obtains $\sum_{x' \neq x}(m_{x'}|\mathcal{M}_x = 0$. Again, the fact that probabilities are nonnegative implies that each term in the sum must vanish, namely $(m_{x'}|\mathcal{M}_x = 0$ for every $x' \neq x$. Now, let $\mathbf{r}$ be a measurement such that $\sum_y r_{xy} = m_x$. Since the measurement $\mathbf{r}$ is compatible with $\mathbf{m}$, Eq. (4) implies $(r_{xy}|\mathcal{M} = (r_{xy}|$ with $\mathcal{M} = \sum_{x'} \mathcal{M}_{x'}$. On the other hand, for $x \neq x'$ the condition $(m_x|\mathcal{M}_{x'} = 0$ implies $(r_{xy}|\mathcal{M}_{x'} = 0$. Hence, we conclude that $(r_{xy}|\mathcal{M}_x = (r_{xy}|\mathcal{M} = (r_{xy}|$ for every x and y. $\qquad\square$

Joint measurability of orthogonal effects. The characterization of sharp instruments, combined with the Less Information-More Sharpness principle, leads directly to the first key result of our work: a construction showing that mutually orthogonal effects can be measured jointly in a single sharp measurement. Precisely, if m_k is orthogonal to m_l for every $k, l \in \{1, \dots, K\}$, we show that there exists a joint sharp measurement $\mathbf{j}$ such that $\{m_k\}_{k=1}^K \subseteq \mathbf{j}$.

Let us see how to construct the joint measurement. Let $\mathbf{m}^{(k)}$ be the sharp measurement that contains the effect m_k. By coarse graining of $\mathbf{m}^{(k)}$ one obtains the binary measurement $\mathbf{m}^{(k)} = \{m_0^{(k)}, m_1^{(k)}\}$, with $m_0^{(k)} := m_k$ and $m_1^{(k)} := u - m_k$. By the Less Information, More Sharpness postulate, $\mathbf{m}^{(k)}$ is sharp. Let $\{\mathcal{M}_0^{(k)}, \mathcal{M}_1^{(k)}\}$ be the corresponding instrument. Now, since m_k and m_l are orthogonal, $\mathbf{m}^{(kl)} = \{m_k, m_l, e - m_k - m_l\}$ must be a valid measurement. Since $\mathbf{m}^{(k)}$ is a coarse-graining of $\mathbf{m}^{(kl)}$, Eq. (6) gives

$$(m_l|\mathcal{M}_1^{(k)} = (m_l|. \qquad (8)$$

Now, consider the following measurement procedure: *i)* perform the first instrument *ii)* if the outcome is 1, then perform the second instrument, *iii)* for every $k < K$, if the outcome of the k-th instrument is 1, perform the $(k+1)$-th instrument. The resulting instrument, denoted by $\{\mathcal{J}_k\}_{i=1}^{K+1}$ consists of the transformations

$$\mathcal{J}_1 := \mathcal{M}_0^{(1)}$$
$$\mathcal{J}_2 := \mathcal{M}_0^{(2)}\mathcal{M}_1^{(1)}$$
$$\mathcal{J}_3 := \mathcal{M}_0^{(3)}\mathcal{M}_1^{(2)}\mathcal{M}_1^{(1)}$$
$$\vdots$$
$$\mathcal{J}_K := \mathcal{M}_0^{(K)}\mathcal{M}_1^{(K-1)}\cdots\mathcal{M}_1^{(1)}$$
$$\mathcal{J}_{K+1} := \mathcal{M}_1^{(K)}\mathcal{M}_1^{(K-1)}\cdots\mathcal{M}_1^{(1)}.$$

The measurement $\mathbf{j} = \{j_k\}_{k=1}^{K+1}$ associated to the instrument $\{\mathcal{J}_k\}_{k=1}^{K+1}$ is the desired joint measurement: indeed, and using Eqs. (2) and (8) we obtain $(j_k| = (e|\mathcal{M}_0^{(k)}\mathcal{M}_1^{(k-1)}\cdots\mathcal{M}_1^{(1)} = (m_k|\mathcal{M}_1^{(k-1)}\cdots\mathcal{M}_1^{(1)} = (m_k|$ for every $k \in \{1, \dots, K\}$. In addition, the measurement $\mathbf{j}$ is sharp. Indeed, if a measurement $\{r_{kl}\}$ is a refinement

of $\mathbf{j}$, i. e. $\sum_l r_{kl} = j_k$ for all k, then $\mathbf{r}$ is also a refinement of the sharp measurement $\mathbf{m}^{(k')}$ for every fixed k'. Hence, one has

$$(r_{kl}|\mathcal{M}_0^{(k)} = (r_{kl}| \qquad \forall k \in \{1, \dots K\}$$
$$(r_{kl}|\mathcal{M}_1^{(k')} = (r_{kl}| \qquad \forall k, k' \in \{1, \dots K\}, k \neq k'.$$

Using this fact and the definition of $\mathcal{J}_k$ it is immediate to obtain the relation $(r_{kl}|\mathcal{J}_k = (r_{kl}|$ for every k, l. Thanks to Eq. (6), this proves that the instrument $\{\mathcal{J}_k\}$ is sharp, and so is the corresponding measurement $\mathbf{j}$. $\qquad\square$

The ability to combine orthogonal effects into a single measurement is a powerful asset. As we already observed, it implies CE at its basic level. In the following we show that it can be used also to obtain the whole CE hierarchy.

Orthogonality of product effects. In a causal theory the information available at a given moment of time can be used to make decisions about the settings of future experiments, thus allowing for adaptive measurements where the choice of setting for a system B depends on the outcome of a measurement on system A. In particular, if $\{m_x\}$ is a sharp measurement on A and $\{n_y^{(x)}\}_{y \in \mathsf{Y}}$ is a sharp measurement for every x, then $\{m_x \otimes n_y^{(x)}\}$ is a legitimate measurement. Now, the Locality of Sharp Measurements implies that $\{m_x \otimes n_y^{(x_0)}\}$ is sharp for every fixed x_0. Since x_0 is arbitrary, this means that each effect $m_x \otimes n_y^{(x)}$ is sharp and that two effects $m_x \otimes n_y^{(x)}$ and $m_{x'} \otimes n_{y'}^{(x')}$ are orthogonal unless $x = x'$ and $y = y'$.

Thanks to this observation, it is easy to see that every level of the CE hierarchy is satisfied. The key is to note that if two strings of outcomes $\mathbf{y}$ and $\mathbf{y}'$ are exclusive, then the corresponding effects $P_\mathbf{y}$ and $P_{\mathbf{y}'}$ are orthogonal. This is clear because, by definition, the effects corresponding to two exclusive strings are of the form $P_\mathbf{y} = m_{y_i} \otimes n$ and $P_{\mathbf{y}'} = m_{y_i'} \otimes n'$ where the effects m_{y_i} and $m_{y_i'}$ are orthogonal and the effects $n = \otimes_{j \neq i} m_{y_j}$ and $n' = \otimes_{j \neq i} m_{y_j}$ are sharp thanks to the Locality of Sharp Measurements. Using our result about product effects, we then have that $P_\mathbf{y}$ and $P_{\mathbf{y}'}$ are orthogonal. Now, a set of mutually exclusive strings E corresponds to a set of mutually orthogonal effects $\{P_\mathbf{y}\}_{\mathbf{y} \in \mathsf{E}}$. Since mutually orthogonal effects can be combined into a joint measurement, the probabilities $p_L(\mathbf{y}) = (P_\mathbf{y}|\rho^{\otimes L})$ obey the bound $\sum_{\mathbf{y} \in \mathsf{E}} p_L(\mathbf{y}) \leq 1$, meaning that the theory satisfies the L-th level of the CE hierarchy for arbitrary L.

Note that the same argument can be used to prove the validity of LO for the probability distributions generated in a scenario where all parties perform sharp measurements. In this scenario, two locally orthogonal events $(\mathbf{x}, \mathbf{y})$ and $(\mathbf{x}', \mathbf{y}')$ correspond to two orthogonal effects $P_\mathbf{y}^{(\mathbf{x})}$ and $P_{\mathbf{y}'}^{(\mathbf{x}')}$, for exactly the same reason mentioned above. Hence, the joint measurability of orthogonal effect implies the bound $\sum_{(\mathbf{x},\mathbf{y}) \in \mathsf{O}} p(\mathbf{x}|\mathbf{y}) \leq 1$ for every set O of locally orthogonal events. In other words, all the prob-

ability distributions generated by sharp measurements obey LO.

Reduction to sharp measurements. While CE applies only to sharp measurements, LO applies to arbitrary measurements. This is because every probability distribution that we can encounter in our theory is a probability distribution generated by sharp measurements. This fact can be seen as follows: combining the Fundamental Sharpness with the Locality of Sharp Measurements, one can show that for every party i and every measurement $\mathbf{m}^{(i,x_i)}$, there exists an ancilla A_i, a state of A_i, call it σ_i, and a sharp measurement $\mathbf{M}^{(i,x_i)}$ such that $\left(m_{y_i}^{(i,x_i)}|\rho_i\right) = \left(M_{y_i}^{(i,x_i)}|\rho_i \otimes \sigma_i\right)$ for every x_i and for every y_i (see the Appendix for the proof). Now, since all measurements that party i can perform can be replaced by sharp measurements by adding an ancilla in a fixed state σ_i, the input-output distribution $p(\mathbf{y}|\mathbf{x})$ generated by arbitrary measurements on the state ρ coincides with the input-output distribution generated by sharp measurements on the state $\rho' = \rho \otimes \sigma_1 \otimes \cdots \otimes \sigma_N$. In other words, at the level of correlations there is no difference between sharp and non-sharp measurements. Thanks to this fact, deriving LO for sharp measurements is equivalent to deriving LO for arbitrary measurements.

ACKNOWLEDGMENTS

[1] A. Einstein, B. Podolsky, and N. Rosen, Phys. Rev. **47**, 777 (1935).
[2] J. S. Bell, *On the Einstein-Podolsky-Rosen Paradox. Physics 1, 195–200 (1964)*, Speakable and Unspeakable in Quantum Mechanics (Cambridge University Press, 1987).
[3] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).
[4] S. Kochen and E. P. Specker, Journal of Mathematics and Mechanics **17**, 59 (1967).
[5] N. D. Mermin, Rev. Mod. Phys. **65**, 803 (1993).
[6] S. Popescu and D. Rohrlich, Foundations of Physics **24**, 379 (1994).
[7] S. Popescu and D. Rohrlich, Foundations of Physics **24**, 379 (1994).
[8] G. Brassard, H. Buhrman, N. Linden, A. A. Méthot, A. Tapp, and F. Unger, Phys. Rev. Lett. **96**, 250401 (2006).
[9] N. Linden, S. Popescu, A. J. Short, and A. Winter, Phys. Rev. Lett. **99**, 180502 (2007).
[10] M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Żukowski, Nature **461**, 1101 (2009).
[11] M. Navascués and H. Wunderlich, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science **466**, 881 (2010).
[12] T. Fritz, A. B. Sainz, R. Augusiak, J. B. Brask, R. Chaves, A. Leverrier, and A. Acín, Nature communications **4**, 2263 (2013).
[13] E. Specker, Dialectica **14**, 239 (1960).
[14] Y.-C. Liang, R. W. Spekkens, and H. M. Wiseman, Physics Reports **506**, 1 (2011).
[15] A. Cabello, Phys. Rev. Lett. **110**, 060402 (2013).
[16] A. Cabello, S. Severini, and A. Winter, Phys. Rev. Lett. **112**, 040401 (2014).

[17] J. Henson, arXiv:1210.5978 (2012).
[18] A. Acín, T. Fritz, A. Leverrier, and A. B. Sainz, arXiv:1212.4084 (2012).
[19] B. Yan, Phys. Rev. Lett. **110**, 260406 (2013).
[20] M. Navascués, Y. Guryanova, M. Hoban, and A. Acín, arXiv:1403.4621 (2014).
[21] B. Coecke, Contemporary physics **51**, 59 (2010).
[22] L. Hardy, arXiv:quant-ph/0101012 (2001).
[23] J. Barrett, Phys. Rev. A **75**, 032304 (2007).
[24] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. A **81**, 062348 (2010).
[25] Electronic Notes in Theoretical Computer Science **270**, 3 (2011).
[26] R. Gallego, L. E. Würflinger, A. Acín, and M. Navascués, Phys. Rev. Lett. **107**, 210403 (2011).
[27] E. Davies, *Quantum theory of open systems* (Academic Press, 1976).
[28] M. Ozawa, Journal of Mathematical Physics **25** (1984).
[29] J. J. Sakurai, *Modern quantum mechanics*, Vol. 1 (1985).
[30] P. Busch, P. Lahti, and P. Mittelstaedt, *The Quantum Theory of Measurement*, Environmental Engineering No. v. 2 (Springer, 1996).
[31] G. Lüders, Annalen der Physik **443**, 322 (1950).
[32] C. Helstrom, Journal of Statistical Physics **1**, 231 (1969).
[33] A. S. Holevo, *Probabilistic and statistical aspects of quantum theory*, Vol. 1 (Springer, 2011).
[34] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. A **84**, 012311 (2011).
[35] L. Hardy, arXiv:1104.2066 (2011).
[36] L. Masanes and M. P. Müller, New Journal of Physics **13**, 063001 (2011).
[37] B. Dakic and C. Brukner, *Quantum theory and beyond: is entanglement special?*, Deep Beauty (Cambridge University Press, 2009).

[38] L. Masanes, M. P. Müller, R. Augusiak, and D. Perez-Garcia, arXiv:1208.0493 (2012).

[39] A. A. Klyachko, M. A. Can, S. Binicioğlu, and A. S. Shumovsky, Phys. Rev. Lett. **101**, 020403 (2008).

[40] M. L. Almeida, J.-D. Bancal, N. Brunner, A. Acín, N. Gisin, and S. Pironio, Phys. Rev. Lett. **104**, 230404 (2010).

[41] C. Berge, *Graphs and hypergraphs*, North-Holland mathematical library (North-Holland Pub. Co., 1973).

[42] M. D'Ariano, AIP Conference Proceedings **844**, 101 (2006).

[43] J. Barrett, Phys. Rev. A **75**, 032304 (2007).

[44] R. W. Spekkens, Phys. Rev. A **75**, 032110 (2007).

[45] B. Schumacher and M. D. Westmoreland, Foundations of Physics **42**, 918 (2012).

[46] J. V. Neumann, *Mathematical Foundations of Quantum Mechanics*, Investigations in physics (Princeton University Press, 1955).

[47] P. A. M. Dirac, *The principles of quantum mechanics*, 27 (Oxford university press, 1981).

[48] A. Holevo, Reports on Mathematical Physics **22**, 385 (1985).

[49] C. Piron, *Foundations of quantum physics*, Mathematical Physics Monograph Series (Benjamin-Cummings Publishing Company, 1976).

[50] C. Piron, Erkenntnis **16**, 397 (1981).

[51] E. G. Beltrametti and G. Cassinelli, *The logic of quantum mechanics*, Vol. 15 (Cambridge University Press, 2010).

[52] C. Pfister and S. Wehner, Nature communications **4**, 1851 (2013).

[53] P. Kurzyński, A. Cabello, and D. Kaszlikowski, Phys. Rev. Lett. **112**, 100401 (2014).

[54] Y. Aharonov, P. G. Bergmann, and J. L. Lebowitz, Phys. Rev. **134**, B1410 (1964).

[55] Y. Aharonov, S. Popescu, and J. Tollaksen, Phys. Today **63**, 27 (2010).

SHARP REALIZATION OF MULTIPLE MEASUREMENTS

Here we show that the Fundamental Sharpness principle implies that for every party i one can find an ancilla and a state of the ancilla, independent on the setting x_i, such that every measurement of party i can be realized as a sharp measurement on the system and the ancilla. This fact is a consequence of the following:

Proposition 1. *Let $\{\mathbf{m}^{(x)}, x \in \mathsf{X}\}$ be a finite set of measurements on system A and let Y_x be the set of outcomes for measurement $\mathbf{m}^{(x)}$. If the Fundamental Sharpness and the Locality of Sharp Instruments hold, then there exists an ancilla B, a state σ of the ancilla, and a sharp measurement $\mathbf{M}^{(x)}$ such that*

$$\left(m_y^{(x)} | \rho \right) = \left(M_y^{(x)} | \rho \otimes \sigma \right) \qquad \forall x \in \mathsf{X}, \forall y \in \mathsf{Y}_x$$

for every state ρ of system A.

Proof. For every setting x, the Fundamental Sharpness principle ensures that there exists an ancilla A_x, a state σ_x of the ancilla, and a sharp measurement $\mathbf{S}^{(x)}$ such that, for every state ρ, one has

$$\left(m_y^{(x)} | \rho \right) = \left(S_y^{(x)} | \rho \otimes \sigma_x \right) \qquad \forall y \in \mathsf{Y}_x,$$

Now, take as ancilla the composite system $B := \bigotimes_x B_x$, define the state $\sigma := \bigotimes_x \sigma_x$, and let $\mathbf{M}^{(x)}$ be the measurement with effects $M_y^{(x)} := S_y^{(x)} \otimes u_{\neg x}$, where $u_{\neg x}$ is the unit effect on all ancillas $A_{x'}$ with $x' \neq x$. Now, $\mathbf{M}^{(x)}$ is the product of the sharp measurement $\mathbf{S}^{(x)}$ with the trivial measurement, which, by definition, is also sharp. Hence, by the Locality of Sharp Measurements, $\mathbf{M}^{(x)}$ is sharp. By construction, one has

$$\left(M_y^{(x)\prime} | \rho \otimes \sigma \right) = \left(S_y^{(x)} | \rho \otimes \sigma \right)$$
$$= \left(m_y^{(x)} | \rho_i \right)$$

for every setting x, every outcome y, and every state ρ. $\qquad \square$