Generalized Gleason theorem and finite amount of information for the context

A. Montina and S. Wolf
Facoltà di Informatica, Università della Svizzera Italiana, Via G. Buffi 13, 6900 Lugano, Switzerland
(Dated: September 26, 2022)

Quantum processes cannot be reduced, in a nontrivial way, to classical processes without specifying the context in the description of a measurement procedure. This requirement is implied by the Kochen-Specker theorem in the outcome-deterministic case and, more generally, by the Gleason theorem. The latter establishes that there is only one non-contextual classical model compatible with quantum theory, the one that trivially identifies the quantum state with the classical state. However, this model requires a breaking of the unitary evolution to account for macroscopic realism. Thus, a causal classical model compatible with the unitary evolution of the quantum state is necessarily contextual at some extent. Inspired by well-known results in quantum communication complexity, we consider a particular class of hidden variable theories by assuming that the amount of information about the measurement context is finite. Aiming at establishing some general features of these theories, we first present a generalized version of the Gleason theorem and provide a simple proof of it. Assuming that Gleason's hypotheses hold only locally for 'small' changes of the measurement procedure, we obtain almost the same conclusion of the original theorem about the functional form of the probability measure. An additional constant and a relaxed property of the 'density operator' are the only two differences from the original result. By this generalization of the Gleason theorem and the assumption of finite information for the context, we prove that the probabilities over three or more outcomes of a projective measurement must be linear functions of the projectors associated with the outcomes, given the information on the context.

I. INTRODUCTION

In the formalism of quantum theory, the possible outcomes of a von Neumann measurement are labeled by projectors. This description provides an operationally exhaustive summary of the whole measurement procedure and contains the complete information that is relevant for distinguishing two events that occur with different probability for some preparation procedure. Furthermore, this labeling is also minimal, that is, it does not distinguish events that always occur with the same probability. Given a measurement, the outcomes are identified by a set of M commuting projectors, say $\{\hat E_1, \hat E_2, \cdots, \hat E_M\}$, with $\sum_{n=1}^M \hat E_n = \mathbb{1}$. The probability of outcome $\hat E_k$, say $\mu(\hat E_k)$, is given by the Born rule

$$\mu(\hat E_k) = \mathrm{Tr}(\hat E_k \hat\rho), \tag{1}$$

where $\hat\rho$ is the density operator, which gives the statistically relevant information about the preparation procedure. The additional information about the measurement procedure that is irrelevant for the computation of $\mu(\hat E_k)$ is referred as *context*. For example, the projectors $\hat E_2, \hat E_3, \ldots, \hat E_M$ are part of the context for the outcome $\hat E_1$, as they are not relevant for computing $\mu(\hat E_1)$.

Although this formalism is operationally exhaustive and minimal, it does not provide a unified description of observed and observing systems. Indeed, two different languages are used for the experimental apparatus and the quantum system under observation. On the one hand the experimental apparatus is described by a purely classical language that specifies for example the position and orientation of beam splitters, mirrors or crystals in a quantum optics experiment. On the other hand the quantum system is indirectly described by the operations performed on the experimental apparatus. The quantum state is not meant as a classical object, such as a field, but it is a mere container of information about the preparation procedure. This formalism is completely silent on the actual state of affairs of each single quantum system. Is it possible to have a unified description that puts experimental apparatus and quantum system on the same footing? Various no-go theorems show that this embedding of quantum processes in the classical framework is not possible without apparently unphysical consequences, such as non-locality [1, 2] and, more generally, contextuality [3]. The latter refers to a dependence of the outcomes on the details of the measurement implementation that are irrelevant for the computation of the quantum probabilities $\mu(\hat E_k)$. Apart from foundational motivations concerning the interpretation of the quantum formalism, a classical embedding of quantum theory is important in the context of quantum information theory, since no-go theorems concerning this embedding make the gap between quantum and classical information more definite. For example, Bell theorem led to the discovery of quantum cryptographic protocols exploiting non-locality as a resource [4]. In quantum communication complexity, classical simulations are relevant for setting a limit on the advantages offered by quantum channels.

Hidden variable (HV) theories [5], also known as ontological theories, are one possible classical reinterpretation of quantum processes. In a HV theory, there is no dichotomy between classical and microscopic quantum world. Any system is always supposed to be in some well-defined classical state, say x, which is an element of a classical space, X. According to the present terminology [6], we will refer to the classical state and the classical space as *ontic state* and *ontological space*, respectively. Given a preparation procedure, the system is set in an ontic state according to some probability distribution that depends on the preparation procedure. When a measurement is performed, the probability of an outcome is conditioned by the ontic state. If the outcome is completely determined, then the ontological theory is said to be *outcome-deterministic*.

The de Broglie-Bohm (dBB) theory is a particular example of outcome-deterministic HV theory. In this case, the quantum state assumes the role of an actual physical field that pilots the dynamics of the particles. Thus, the ontic state x is identified with the wave-function and the positions of the particles. Another example is given by the Beltrametti-Bugajski (BB) model. Differently from dBB theory, the ontic state is identified with only the quantum state, which is not supplemented by any additional variable. Furthermore, the BB model is not outcome-deterministic. In more general HV theories, the ontic state does not necessarily contain the full information about the quantum state, which is instead encoded into the statistical behavior of many identically prepared realizations. We call an ontological theory *trivial* if, for any measurement, the outcome probabilities, given an ontic state x, are equal to the quantum probabilities, given some quantum state $|\psi\rangle$. In this case, x can be identified with $|\psi\rangle$. According to this definition, the BB model is a trivial ontological theory. Any trivial HV theory is essentially equivalent to the BB model. The dBB theory, being deterministic, is a counterexample of nontrivial HV theory.

In their seminal article [3], Kochen and Specker showed that any outcome-deterministic ontological theory is measurement-contextual. In other words, the minimal labeling of an event with a projector is not sufficient to describe consistently a measurement procedure. In Ref. [6], it was pointed out that the outcome determinism is a necessary condition for inferring the measurement contextuality. Indeed the BB model is an example of measurement-noncontextual ontological theory, which is not outcome-deterministic. In fact the BB model is the only noncontextual ontological theory [7]. Equivalently, quantum mechanics is essentially the only theory that employs the minimal labeling in the description of the events. In Ref. [7], we argued that the BB model, being a trivial HV theory, is not sufficient for introducing realism in the quantum phenomena, unless the unitarity of the evolution is broken. Thus, we concluded that the measurement contextuality should be introduced to some extent. What is the minimal amount of required information about the context? It is a well-known result of quantum information that a finite amount of classical communication is sufficient to reproduce classically the quantum correlations in an Einstein-Podolsky-Rosen (EPR) experiment [11]. This kind of correlations provides a particular example of measurement contextuality [5]. Inspired by this result of quantum information, we consider a particular class of HV theories by assuming that the amount of relevant information about the measurement context is always finite. Using this hypothesis and a generalized version of the Gleason theorem, we prove that the probability of an event must be a linear function of the projector associated with the event, given the information on the context. This result provides an example illustrating the relevance of the generalized Gleason theorem. As remarked in the conclusion, this theorem can turn to be useful for solving some general questions in quantum information.

The paper is organized as follows. In Sec. II we introduce the hypothesis that the amount of relevant information about the context in a HV theory is finite. By this hypothesis, we show that the ontological theory is somehow noncontextual for 'small' changes of the measurement procedure. In Sec. III, we prove the generalized Gleason theorem. Assuming that Gleason's hypotheses hold only locally for 'small' changes of the measurement procedure, we obtain almost the same conclusion of the original theorem about the functional form of the probability measure. An additional constant and a relaxed property of the 'density operator' are the two only differences from the original result. The proof is much simpler than Gleason's proof. Furthermore the Gleason theorem can be derived as a corollary from this generalization. In Sec. IV, we derive the general form of the probability of an event in a HV theory by using the hypothesis in Sec. II and the generalized Gleason theorem. Finally, the conclusion and the perspectives are drawn.

II. FINITE AMOUNT OF INFORMATION ABOUT THE CONTEXT

Let us first introduce the framework of an ontological theory. For convenience, in the following we will associate projective measurements with ordered M-tuples of projectors. A quantum system is described by an ontic state, x, which is an element of an ontological space, X. When the quantum system is prepared in a quantum state $|\psi\rangle$, its ontic state is set according to a probability distribution $\rho(x|\psi,\eta)$ which depends on $|\psi\rangle$ and, possibly, an additional parameter, η, representing the preparation context. Thus, we have the mapping

$$(|\psi\rangle,\eta)\rightarrow\rho(x|\psi,\eta), \tag{2}$$

where $(|\psi\rangle,\eta)$ represents the preparation procedure. When a measurement $\mathcal{M}=(\hat E_1,\hat E_2,\cdots,\hat E_M)$ is performed, the probability of an outcome $\hat E_k$, is conditioned by the value of x. Differently from the quantum formalism, in general the probability also depends on the whole set $\mathcal{M}$, not just $\hat E_k$. The other projectors give the measurement context for the event $\hat E_k$. This dependence is not the only possible kind of contextuality. We denote by τ the additional context. Thus, a measurement procedure is specified by the pair $(\mathcal{M},\tau)$, which is associated with a conditional probability of having outcome $\hat E_k$ given x, that is,

$$(\mathcal{M},\tau)\rightarrow P(\hat E_k|x,\mathcal{M},\tau). \tag{3}$$

The set $\mathcal{M}$ is *complete*, that is, the sum of the projectors in $\mathcal{M}$ is the identity operator,

$$\sum_k \hat{E}_k = \hat{\mathbb{1}}. \tag{4}$$

The probability distribution P satisfies the normalization equation

$$\sum_{k=1}^{M} P(\hat{E}_k|x,\mathcal{M},\tau) = 1. \tag{5}$$

The ontological model reproduces a process of state preparation and subsequent measurement if the equality

$$\int dx P(\hat{E}_k|x,\mathcal{M},\tau)\rho(x|\psi,\eta) = \langle\psi|\hat{E}_k|\psi\rangle \tag{6}$$

is satisfied, where the integral is defined according to some measure on X. In quantum communication complexity, this equation describes the simulation of a noiseless quantum channel with subsequent projective measurement.

Hereafter, we consider the class of ontological theories for which the amount of relevant information about the measurement context is finite. Thus, we assume that the conditional probability in Eq. (3) takes the form

$$P(\hat{E}_k|x,\mathcal{M},\tau) = \sum_n \mu(\hat{E}_k|x,n) P_c(n|x,\mathcal{M},\tau), \tag{7}$$

If the worst-case amount of information is finite, then the sum in Eq. (7) is over a finite number of elements. More generally, we only assume that the index n is discrete and the summation $\sum_n P_c(n|x,\mathcal{M},\tau)$ converges to 1. Note that the probability distribution P_c explicitly depends on the ontic state, that is, the information about the context generally depends on the value x in each single realization. Indeed, in the EPR scenario, this dependence is necessary if the context is summarized by a finite amount of information [12]. Also note that there is a redundancy in the definition of $\mu(\hat{E}_k|x,n)$, since the index n can contain some information about $\hat{E}_k$. This implies that there could be a conflict between the value of n and $\hat{E}_k$. For example, if the probability $P_c(n'|x,\mathcal{M},\tau)$ is equal to zero for some n' and for every $(\mathcal{M},\tau)$ such that E_k is equal to some E'_k, then $\mu(\hat{E}'_k|x,n')$ is left indeterminate. Indeed, denoting by $P(\hat{E}_k,n|x,\mathcal{M},\tau)$ the joint probability of $\hat{E}_k$ and n, we have that

$$\mu(\hat{E}'_k|x,n') = \frac{P(\hat{E}'_k,n'|x,\mathcal{M},\tau)}{P_c(n'|x,\mathcal{M},\tau)} \text{ for } \hat{E}_k = \hat{E}'_k$$

and both the numerators and denominators are zero. This point is important for correctly deriving the properties of the conditional probability $\mu(\hat{E}_k|x,n)$. In particular, the normalization condition $\sum_k \mu(\hat{E}_k|x,n) = 1$ is required only if n is consistent with $\hat{E}_k$ for every k, that is, if $P_c(n|x,\mathcal{M},\tau) \neq 0$ for some τ. Similarly, the non-negativity condition $\mu(\hat{E}_k|x,n) \geq 0$ holds if $\hat{E}_k$ and n are consistent.

To state concisely the normalization condition for $\mu(\hat{E}_k|x,n)$, let us introduce some set definition. We denote by Ω the set containing all the elements $\mathcal{M}$. This set, endowed with a Riemannian metric, is a Riemannian manifold. It has disjoint subsets and each subset contains elements whose projectors $\hat{E}_k$ have fixed rank.

Definition 1 *$\Omega_n(x)$ is the the largest subset of Ω such that, for every $\mathcal{M} \in \Omega_n(x)$, $P_c(n|x,\mathcal{M},\tau) \neq 0$ for some τ.*

In other words, the set $\Omega_n(x)$ contains all the measurements that are consistent with the context index n. The sets $\Omega_n(x)$ cover the set Ω, that is,

$$\cup_n \Omega_n(x) = \Omega. \tag{8}$$

Given this definition, the normalization of the conditional probability $\mu(\hat{E}_k|x,n)$ and its non-negativity can be stated as follows.

$$\begin{array}{l} (\hat{E}_1,\cdots,\hat{E}_M) \in \Omega_n(x) \Rightarrow \\ \left\{ \begin{array}{l} \sum_{k=1}^{M} \mu(\hat{E}_k|x,n) = 1 \\ \mu(\hat{E}_k|x,n) \geq 0 \quad k \in \{1,\dots,M\} \end{array} \right. . \end{array} \tag{9}$$

This property, the normalization of $P_c(n|x,\mathcal{M},\tau)$ and its non-negativity guarantee that the probability distribution $P(\hat{E}_k|x,\mathcal{M},\tau)$ defined by Eq. (7) is normalized and non-negative.

Since the family of sets $\Omega_n(x)$ is countable, we can replace these sets with open sets by removing zero-measure boundaries. The resulting family is identical to the original one up to a negligible zero-measure set of measurements. Thus, we can just assume that the sets $\Omega_n(x)$ are open without loss of generality. Furthermore, we can assume that they are connected. Indeed, if the sets are not connected, we first can write them as union of connected sets,

$$\Omega_n(x) = \cup_k \Omega_{n,k}(x) \tag{10}$$

and replace the probability distribution $P_c(n|x,\mathcal{M},\tau)$ with

$$P_c(n,k|x,\mathcal{M},\tau) \equiv P_c(n|x,\mathcal{M},\tau)\,\delta[\mathcal{M} \in \Omega_{n,k}(x)], \tag{11}$$

where $\delta[\text{true}] = 1$ and $\delta[\text{false}] = 0$. Then, we can rename the pair (n,k) by using only one discrete index n. In this way, we obtain a new model and new corresponding sets $\Omega_n(x)$ that are open and connected. Thus, we can assume that the sets $\Omega_n(x)$ satisfy the following.

Property 1 *The sets $\Omega_n(x)$ are open and connected for every n and every x.*

A measurement $(\hat{E}_1+\hat{E}_2,\dots,\hat{E}_M)\equiv\mathcal{M}_c$ can be implemented as the coarse graining of the measurement $(\hat{E}_1,\hat{E}_2,\cdots,\hat{E}_M)=\mathcal{M}$. Thus, we have the inference

$$\mathcal{M}\in\Omega_n(x)\Rightarrow\mathcal{M}_c\in\Omega_n(x). \tag{12}$$

Indeed, if $\mathcal{M}\in\Omega_n(x)$, then there is a context τ such that $P_c(n|x,\mathcal{M}_c,\tau)\neq 0$. In general, the opposite inference is not true, that is,

$$\mathcal{M}_c\in\Omega_n(x)\not\Rightarrow\mathcal{M}\in\Omega_n(x). \tag{13}$$

Indeed, the measurement $\mathcal{M}_c$ could be implemented without involving a coarse graining of $\mathcal{M}$. Thus, the two measurements could be associated with different values of the context index n.

For the following discussions, it is useful to define the operators $\mathcal{P}_k$.

Definition 2 *Let S be a set of ordered m-tuples. $\mathcal{P}_k S$ with $k\in\{1,\dots,m\}$ is a set such that an element p is in $\mathcal{P}_k S$ if and only if there is an M-tuple b in S whose k-th component is equal to p.*

Thus, the operator $\mathcal{P}_k$ is a kind of Cartesian projector. Similarly, let us define the operators $\mathcal{P}_{kl}$.

Definition 3 *Let S be a set of ordered m-tuples. $\mathcal{P}_{kl}S$ with $k,l\in\{1,\dots,m\}$ and $k\neq l$ is a set of pairs such that an element (p_1,p_2) is in $\mathcal{P}_{k,l}S$ if and only if there is an M-tuple b in S whose k-th and l-th components are equal to p_1 and p_2, respectively.*

By inference (12), we have that property (9) is equivalent to the following ones

$$\hat{E}\in\mathcal{P}_k\Omega_n(x)\Rightarrow\mu(\hat{E}|x,n)\geq 0, \tag{14}$$

$$\begin{gathered}(\hat{E}_1,\hat{E}_2)\in\mathcal{P}_{kl}\Omega_n(x)\Rightarrow\\ \mu(\hat{E}_1|x,n)+\mu(\hat{E}_2|x,n)=\mu(\hat{E}_1+\hat{E}_2|x,n),\end{gathered} \tag{15}$$

$$\mu(\hat{\mathbb{1}})=1. \tag{16}$$

III. GENERALIZED GLEASON THEOREM

Before introducing the generalized Gleason theorem, let us briefly review the original theorem [8]. In the axiomatic formulation of quantum mechanics, each outcome is labeled by a projector $\hat{E}$ and the probability of $\hat{E}$, say $\mu(\hat{E})$, is given by the Born rule

$$\mu(\hat{E})=\mathrm{tr}(\hat{E}\hat{\rho}), \tag{17}$$

where $\hat{\rho}$ is the density operator representing the state of the quantum system. This measure satisfies the two properties

$$\mu(\hat{E})\geq 0, \tag{18}$$

$$\sum_{i=1}^{M}\mu(\hat{E}_i)=1, \tag{19}$$

where $\{\hat{E}_1,\cdots,\hat{E}_M\}$ is any complete set of commuting projectors (so that $\sum_{k=1}^M\hat{E}_k=\hat{\mathbb{1}}$). Eq. (19) is equivalent to the following conditions,

$$\mu(\hat{\mathbb{1}})=1, \tag{20}$$

$$\begin{gathered}\text{for every pair }\{\hat{E}_1,\hat{E}_2\}\text{ of commuting projectors}\\ \Rightarrow\mu(\hat{E}_1)+\mu(\hat{E}_2)=\mu(\hat{E}_1+\hat{E}_2).\end{gathered} \tag{21}$$

Provided that the Hilbert space dimension is greater than 2, Gleason's theorem states that any measure with properties (18) and (19) [or, equivalently, properties (18,20,21)] has the form (17), where $\hat{\rho}$ is a non-negative operator with trace one. This result provides a way for reducing the axiomatic basis of quantum mechanics. Indeed, it shows that the Born rule can be inferred by the assumption that every outcome is associated with a projector and every complete set of commuting projectors represents a complete set of measurement outcomes.

Theorem 1 (Gleason's theorem) *For a Hilbert space of dimension greater than 2, a measure $\mu(\hat{E})$ that satisfies properties (18) and (19) has the form $\mu(\hat{E})=tr(\hat{E}\hat{\rho})$, where $\hat{\rho}$ is a Hermitian non-negative definite matrix with trace equal to 1. Equivalently, the same conclusion holds if properties (18,20,21) are satisfied.*

Now, we present a generalization that has weaker hypotheses than Gleason's theorem and the almost identical conclusion. It requires that the Gleason hypotheses hold locally in some open subset of Ω. We only introduce the additional hypothesis that $\mu(\hat{E})$ is a generalized function (mathematical distribution) [9] for which the derivatives are well-defined in the domain of μ. Indeed, this can be considered the only case that is physically relevant. It is worth stressing that we are not assuming the stronger hypothesis of differentiability. Indeed, a piecewise differentiable function with discontinuities along some zero-measure subset is an example of mathematical distribution. More generally, integrable functions on compact sets are physically relevant examples of distributions. A distribution is formally defined as a functional from a set of test functions to $\mathbb{R}$. With an abuse of notation, we will represent distributions as conventional functions. Our hypothesis on the function μ is complementary to the non-negativity hypothesis used by Gleason [8], which is not required by the generalized Gleason theorem. The latter property, in the original theorem, rules out highly discontinuous unbounded functions satisfying the addition rule in inference (21).

Theorem 2 (Generalization of Gleason's theorem I) *Let O be a connected open set of complete M-tuples of commuting projectors, say $(\hat{E}_1,\dots,\hat{E}_M)$, with $M>2$. Let $\mu(\hat{E})$ be a generalized function whose derivatives are well-defined in $\cup_i\mathcal{P}_iO$. If the equality*

$$\sum_i\mu(\hat{E}_i)=1 \tag{22}$$

is satisfied for every M-tuple in O, then there is an Hermitian operator $\hat\eta$ and real numbers $K_1,\dots,K_M$ such that

$$\hat E\in\mathcal{P}_iO\Rightarrow\mu(\hat E)=tr(\hat\eta\hat E)+K_i. \tag{23}$$

If the intersection of $\mathcal{P}_iO$ and $\mathcal{P}_jO$ is not empty, then $K_i=K_j$.

The theorem can be equivalently stated as follows.

Theorem 3 (Generalization of Gleason's theorem II) *Let O be a connected open set of incomplete pairs of commuting projectors. Let $\mu(\hat E)$ be a distribution whose derivatives are well-defined in $\mathcal{P}_1O\cup\mathcal{P}_2O$. If the property*

$$\mu(\hat E_1)+\mu(\hat E_2)=\mu(\hat E_1+\hat E_2), \tag{24}$$

is satisfied for every pair $(\hat E_1,\hat E_2)\in O$, then there is a Hermitian operator $\hat\eta$ such that

$$\hat E\in\mathcal{P}_iO\Rightarrow\mu(\hat E)=tr(\hat\eta\hat E)+K_i\equiv tr(\hat\eta_i\hat E), \tag{25}$$

where $\hat\eta_i=\hat\eta+(K_i/r_i)\hat{\mathbb{1}}$, r_i being the rank of $\hat E\in\mathcal{P}_iO$. If the intersection of $\mathcal{P}_1O$ and $\mathcal{P}_2O$ is not empty, then $K_1=K_2$ and $r_1=r_2$, so that

$$\hat E\in\mathcal{P}_iO\Rightarrow\mu(\hat E)=tr(\hat\eta\hat E) \tag{26}$$

for some Hermitian operator $\hat\eta$.

It is worth to remark that the operator $\hat\eta$ in Eq. (25) does not depend on the index i. Note that, if the additional hypothesis $\mu(\hat E_i)\ge 0$ is added, the density operator $\hat\eta_i=\hat\eta+(K_i/r_i)\hat{\mathbb{1}}$ is not necessarily non-negative defined. Indeed, the function $\text{tr}(\hat\eta_i\hat E)$ must be positive only in a subset of projectors. The two theorems are equivalent. Indeed, it is clear that Theorem 3 implies Theorem 2. The other inference comes by taking $\hat E_3\equiv\hat{\mathbb{1}}-\hat E_1-\hat E_2$ and $\mu(\hat E_3)\equiv 1-\mu(\hat E_1)-\mu(\hat E_2)$ for $(\hat E_1,\hat E_2)\in O$.

To prove Theorem 3, we first consider projections onto one-dimensional spaces. Thus, we assume that $\hat E_i\equiv\vec\phi_i\vec\phi_i^\dagger$ for $i=1,2$, where $\vec\phi_1$ and $\vec\phi_2$ are two unit orthogonal column vectors. We denote by $\phi_{i;k}$ the components of $\vec\phi_i$ and define the function

$$f(\vec\phi_i)\equiv\mu(\vec\phi_i\vec\phi_i^\dagger),$$

which is called by Gleason *frame function*. Under the restriction of rank-1 projectors, Theorem 3 takes the form of the following.

Lemma 1 *Given a Hilbert space of dimension larger than 2, let $\mathcal{O}$ be an open set of ordered pairs of orthogonal vectors. If, for any pair $(\vec\phi_1,\vec\phi_2)\in\mathcal{O}$, the frame function satisfies the properties*

$$f(\vec\phi_i)\ge 0, \tag{27}$$

$$f(\vec\phi_1)+f(\vec\phi_2)=\mu(\vec\phi_1\vec\phi_1^\dagger+\vec\phi_2\vec\phi_2^\dagger), \tag{28}$$

then the third-order derivatives of $f(\vec\phi)$ with respect to $\vec\phi$ are equal to zero in $\mathcal{P}_1\mathcal{O}\cup\mathcal{P}_2\mathcal{O}$. In particular, if $\mathcal{O}$ is connected, then there is a Hermitian operator $\hat\eta$ and two constants K_1 and K_2 such that

$$\vec\phi\in\mathcal{P}_i\mathcal{O}\Rightarrow f(\vec\phi)=tr(\hat\eta\vec\phi\vec\phi^\dagger)+K_i \tag{29}$$

for $i=1,2$.

Equation (28) states that the sum $f(\vec\phi_1)+f(\vec\phi_2)$ depends only on the subspace spanned by the vectors $\vec\phi_1$ and $\vec\phi_2$. Thus, if the pair $(\vec\chi_1,\vec\chi_2)$ is in $\mathcal{O}$ and the vectors $\vec\chi_i$ are linear combinations of $\vec\phi_i$, then $f(\vec\phi_1)+f(\vec\phi_2)=f(\vec\chi_1)+f(\vec\chi_2)$.

Note that $f(\vec\phi)$ is defined on the unit sphere. As we will see, it is useful to expand the domain of f to the whole vector space and introduce the radial constraint

$$\vec\phi\cdot\frac{\partial f}{\partial\vec\phi}+\vec\phi^*\cdot\frac{\partial f}{\partial\vec\phi^*}=2f. \tag{30}$$

It is always possible to expand the domain and satisfy this constraint with a suitable choice of the radial behavior of f. Indeed, given a function $f(\vec\phi)$ on the unit sphere, the function $f\left(\frac{\vec\phi}{|\vec\phi|}\right)|\vec\phi|^2$ on the vector space is equal to $f(\vec\phi)$ on the unit sphere and satisfies Eq. (30).

Proof of Lemma 1. The main task is to prove that

$$\begin{array}{l}\frac{\partial^3 f(\vec\phi)}{\partial\phi_i\partial\phi_j\partial\phi_k}=0,\\ \frac{\partial^3 f(\vec\phi)}{\partial\phi_i\partial\phi_j\partial\phi_k^*}=0\end{array} \tag{31}$$

and their complex conjugations. For this purpose, it is sufficient to prove the real version of these equalities for a real three-dimensional space.

For any tern $\{i_1,i_2,i_3\}$ of integers such that $i_1\ne i_2,i_3$ and $i_2\ne i_3$, we write the components $\phi_{1;i_k}$ and $\phi_{2;i_k}$ in the form

$$\phi_{1;i_k}\equiv v_k e^{i\varphi_k},\phi_{2;i_k}\equiv w_k e^{i\varphi_k}\quad k\in\{1,2,3\}, \tag{32}$$

where v_k and w_k are components of two orthogonal three-dimensional real vectors, $\vec v$ and $\vec w$ respectively. The task is reduced to prove that

$$\frac{\partial^3 f(\vec v)}{\partial v_i\partial v_j\partial v_k}=0,\ \text{for any } i,j,k\in\{1,2,3\}. \tag{33}$$

The generator of a three-dimensional rotation of $\vec v$ and $\vec w$ is

$$\mathcal{R}(\vec a)=\vec a\cdot\left(\vec v\wedge\frac{\partial}{\partial\vec v}+\vec w\wedge\frac{\partial}{\partial\vec w}\right) \tag{34}$$

where $\vec a$ gives the rotation axis. For a rotation in the plane spanned by the orthogonal vectors $\vec v$ and $\vec w$, we have that $\vec a=\vec v\wedge\vec w$. From Eq. (28), we have that $\mathcal{R}(\vec v\wedge\vec w)[f(\vec v)+f(\vec w)]=0$, that is,

$$\left(\vec w\cdot\frac{\partial}{\partial\vec v}-\vec v\cdot\frac{\partial}{\partial\vec w}\right)[f(\vec v)+f(\vec w)]=0 \tag{35}$$

for every pair of orthogonal vectors $\vec{v}$ and $\vec{w}$ in the domain of definition of f. The generators of the rotations around $\vec{w}$ and $\vec{v}$ are $(\vec{v}\wedge\vec{w})\cdot\frac{\partial}{\partial\vec{v}}$ and $(\vec{v}\wedge\vec{w})\cdot\frac{\partial}{\partial\vec{w}}$, respectively. Applying these operators to both sides of Eq. (35), we obtain the two equations

$$\sum_{ij}(\vec{v}\wedge\vec{w})_i w_j \frac{\partial^2 f(\vec{v})}{\partial v_i \partial v_j} = (\vec{v}\wedge\vec{w})\cdot\frac{\partial}{\partial\vec{w}} f(\vec{w}), \tag{36}$$

$$\sum_{ij}(\vec{v}\wedge\vec{w})_i v_j \frac{\partial^2 f(\vec{w})}{\partial w_i \partial w_j} = (\vec{v}\wedge\vec{w})\cdot\frac{\partial}{\partial\vec{v}} f(\vec{v}). \tag{37}$$

Then, we apply again the operator $\vec{w}\cdot\frac{\partial}{\partial\vec{v}} - \vec{v}\cdot\frac{\partial}{\partial\vec{w}}$ to both sides of Eq. (36) and obtain

$$\begin{array}{c}\sum_{ij}(\vec{v}\wedge\vec{w})_i\left(v_j\frac{\partial^2 f(\vec{v})}{\partial v_i\partial v_j} - w_j\sum_k w_k\frac{\partial^3 f(\vec{v})}{\partial v_i\partial v_j\partial v_k}\right) = \\ \sum_{ij}(\vec{v}\wedge\vec{w})_i v_j\cdot\frac{\partial^2 f(\vec{w})}{\partial w_i\partial w_j}.\end{array} \tag{38}$$

Thus, the left-hand side of this equation is equal to the right-hand side of Eq. (37), that is,

$$\begin{array}{c}\sum_{ijk}(\vec{v}\wedge\vec{w})_i w_j w_k\frac{\partial^3 f(\vec{v})}{\partial v_i\partial v_j\partial v_k} = \\ (\vec{v}\wedge\vec{w})\cdot\frac{\partial}{\partial\vec{v}}\left[\vec{v}\cdot\frac{\partial f(\vec{v})}{\partial\vec{v}} - 2f(\vec{v})\right].\end{array} \tag{39}$$

From Eq. (30) we have that

$$\vec{v}\cdot\frac{\partial f(\vec{v})}{\partial\vec{v}} - 2f(\vec{v}) = 0. \tag{40}$$

Thus, Eqs. (39) and (40) imply that

$$\sum_{ijk} u_i w_j w_k \frac{\partial^3 f(\vec{v})}{\partial v_i\partial v_j\partial v_k} = 0 \tag{41}$$

for every tern $\{\vec{u},\vec{v},\vec{w}\}$ of orthogonal vectors. This implies that

$$\sum_{ijk} w_i w_j w_k \frac{\partial^3 f(\vec{v})}{\partial v_i\partial v_j\partial v_k} = 0 \tag{42}$$

for every pair $\{\vec{v},\vec{w}\}$ of orthogonal vectors. Indeed, this equation can be derived from Eq. (41) by considering the two pairs of orthogonal vectors $(\vec{u},\vec{w}) = (\vec{u}'\mp\vec{w}', \vec{u}'\pm\vec{w}')$, where $\vec{u}'$ and $\vec{w}'$ are vectors orthogonal to $\vec{v}$ and with $|\vec{u}'| = |\vec{w}'|$. These two cases and the equations

$$\sum_{ijk} u'_i w'_j w'_k \frac{\partial^3 f(\vec{v})}{\partial v_i\partial v_j\partial v_k} = \sum_{ijk} u'_i u'_j w'_k \frac{\partial^3 f(\vec{v})}{\partial v_i\partial v_j\partial v_k} = 0$$

give the two equations $(w'_i w'_j w'_k \pm u'_i u'_j u'_k)\frac{\partial^3 f}{\partial v_i\partial v_j\partial v_k} = 0$, which imply Eq. (42). Thus, every third-order derivative in the subspace orthogonal to $\vec{v}$ is equal to zero. Furthermore, from Eq. (40) we have that

$$\vec{v}\cdot\frac{\partial^3 f(\vec{v})}{\partial\vec{v}\partial v_i\partial v_j} = 0 \tag{43}$$

for every $i,j\in\{1,2,3\}$. This implies that every third-order derivative is zero and, thus, Eq. (33) is satisfied. Identical equations hold for $f(\vec{w})$. Since Eq. (33) holds for any real three-dimensional subspace of a complex Hilbert space, also Eqs. (31) and their complex conjugations hold. Thus, the function $f(\vec{\phi}_i)$ must be quadratic in $\vec{\phi}_i$. Since the frame function $f(\vec{\phi}_i)$ is equal to $\mu(\vec{\phi}_i\vec{\phi}_i^\dagger)$, the linear terms and the terms in $\phi_{i;k}\phi_{i;l}$ and $\phi^*_{i;k}\phi^*_{i;l}$ are equal to zero. In particular, if $\mathcal{O}$ is connected, then

$$f(\vec{\phi}) = tr(\hat{\eta}\vec{\phi}\vec{\phi}^\dagger) + K_i \tag{44}$$

for $\vec{\phi}\in\mathcal{P}_i\mathcal{O}$ and $i\in\{1,2\}$. The lemma is proved. $\square$

Proof of Theorem 3. We can decompose the two projectors $\hat{E}_1$ and $\hat{E}_2$ into rank-1 commuting projectors, say $F_i^{(k)}$,

$$\hat{E}_1 = \sum_{k=1}^{r_1}\hat{F}_1^{(k)}, \;\; \hat{E}_2 = \sum_{k=1}^{r_2}\hat{F}_2^{(k)}, \tag{45}$$

where r_i is the rank of $\hat{E}_i$. Since the pair $(\hat{E}_1,\hat{E}_2)$ is not complete, we have that $r_1+r_2<N$, where N is the dimension of the Hilbert space. Let us denote by $\vec{\phi}_i^k$ the vectors such that $\hat{F}_i^{(k)} = \vec{\phi}_i^k(\vec{\phi}_i^k)^\dagger$. By Lemma (1), we have that the function μ must be linear in $\hat{F}_i^{(k)}$ in each connected set for any decomposition of $\hat{E}_1$ and $\hat{E}_2$. For example, keeping $\hat{F}_1^{(2)},\cdots,\hat{F}_1^{(r_1)}$ and $\hat{F}_2^{(2)},\cdots,\hat{F}_2^{(r_2)}$ constant. The orthogonal complement, say $\mathcal{H}_\perp$, of the subspace spanned by $\vec{\phi}_1^2,\ldots,\vec{\phi}_1^{r_1}$ and $\vec{\phi}_2^2,\ldots,\vec{\phi}_2^{r_2}$ is a vector subspace of dimension equal to $N-r_1-r_2+2>2$. We denote by Π the set of pairs of orthogonal vectors in $\mathcal{H}_\perp$. The set of pairs $(\vec{\phi}_1^1,\vec{\phi}_2^1)$ such that $(\hat{E}_1,\hat{E}_2)\in O$ is an open set of Π. Thus, by Lemma 1, the third-order derivatives of μ with respect to $\vec{\phi}_1^1$ and $\vec{\phi}_2^1$ are equal to zero. This is true for every decomposition of $\hat{E}_i$ into rank-1 projectors. This implies that μ is linear in $\hat{E}_1$ and $\hat{E}_2$ and has the form

$$\mu(\hat{E}_i) = \mathrm{tr}(\hat{\eta}\hat{E}_i) + K_i \tag{46}$$

for any $\hat{E}_i\in\mathcal{P}_iO$. If the intersection of $\mathcal{P}_1O$ and $\mathcal{P}_2O$ is not empty, it is trivial that $K_1=K_2$ and $r_1=r_2$. $\square$

Gleason's theorem is a trivial consequence of theorem 3. Indeed, if the set O in the theorem statement is the whole set of pairs of commuting projectors with fixed rank r, then the coefficients K_i in Eq. (25) are independent of i, $K_i=K$. Let us define $\hat{\rho}\equiv\hat{\eta}+Kr^{-1}\hat{\mathbb{1}}$, Eq. (25) gives

$$\mu(\hat{E}) = tr(\hat{\rho}\hat{E}). \tag{47}$$

The non-negativity of $\mu(\hat{E})$ and the equality $\mu(\hat{\mathbb{1}})=1$ imply that $\hat{\rho}$ is positive semidefinite and with trace equal to 1.

IV. FUNCTIONAL FORM OF THE OUTCOME PROBABILITY IN A HV THEORY

The generalization of the Gleason theorem has an obvious consequence for the functional form of the conditional probability $\mu(\hat{E}|x,n)$ defined in Sec. II. Let us remind that this function satisfies the three conditions

$$\begin{gathered}\hat{E}\in\mathcal{P}_k\Omega_n(x)\Rightarrow\mu(\hat{E}|x,n)\ge 0,\\ (\hat{E}_1,\hat{E}_2)\in\mathcal{P}_{kl}\Omega_n(x)\Rightarrow\\ \mu(\hat{E}_1|x,n)+\mu(\hat{E}_2|x,n)=\mu(\hat{E}_1+\hat{E}_2|x,n),\\ \mu(\hat{\mathbb{1}}|x,n)=1.\end{gathered}$$

stated in Eqs. (14,15,16). In particular, the second hypothesis is identical to that used for the generalized Gleason theorem. Indeed, according to property 1, the set $\mathcal{P}_{kl}\Omega_n(x)$ is a connected open set apart from a negligible boundary. Thus, the generalized Gleason theorem implies that

$$\hat{E}\in\mathcal{P}_k\bar{\Omega}_n(x)\Rightarrow\mu(\hat{E}|x,n)=\text{tr}\left[\hat{\eta}(x,n)\hat{E}\right]+K(x,n), \tag{48}$$

where $\bar{\Omega}_n(x)$ is the subset of $\Omega_n(x)$ containing measurements with three or more outcomes. Inferences (9) also imply that

$$\begin{gathered}\{\hat{E}_1,\cdots,\hat{E}_M\}\in\bar{\Omega}_n(x)\Rightarrow\\ \sum_{k=1}^M\left\{\text{tr}[\hat{E}_k\hat{\eta}(x,n)]+K(x,n)\right\}=1,\end{gathered} \tag{49}$$

$$\begin{gathered}\{\hat{E}_1,\cdots,\hat{E}_M\}\in\bar{\Omega}_n(x)\Rightarrow\\ \text{tr}[\hat{E}_k\hat{\eta}(x,n)]+K(x,n)\ge 0.\end{gathered} \tag{50}$$

Note that an outcome-deterministic theory is compatible with the equations that we have derived from the generalized Gleason theorem. Indeed, the ontological theory is outcome-deterministic if, for example, $\hat{\eta}(x,n)=0$ and $K(x,n)$ is identically equal to 0 or 1 where $P_c(n|x,\mathcal{M},\tau)\rho(x|\psi,\eta)$ is different from zero.

In fact, we have not proved that there is an ontological model such that the context information is finite, we have only shown that, if it exists, then it must have some general structure. However, an approximate classical protocol simulating entanglement and quantum channels, reported in Ref. [10], would suggest that such a model exists. In this section, we have assumed that the measurement is performed all at once. This justifies why the context for the event $\hat{E}_k$, in general, depends on the whole set M of projectors. Given multiple commuting measurements, causality imposes some further constraints. Suppose that two-outcome measurements are performed consecutively with outcomes $\{\hat{E}_1,\hat{\mathbb{1}}-\hat{E}_1\}$, $\{\hat{E}_2,\hat{\mathbb{1}}-\hat{E}_2\}$,..., $\{\hat{E}_M,\hat{\mathbb{1}}-\hat{E}_M\}$, where $\hat{E}_k$ are commuting projectors. Under the hypothesis of causality, the outcome of a measurement cannot be influenced by future measurements. Thus, we can rearrange the context index n as an M-tuple of indices $(n_1,\dots,n_M)\equiv\vec{n}$ so that the conditional probability μ of $\hat{E}_k$ given $\vec{n}$ depends only on the first m indices, that is,

$$\begin{aligned}&\mu(\hat{E}_1|x,\vec{n})=\mu(\hat{E}_1|x,n_1)\\ &\dots\\ &\mu(\hat{E}_M|x,\vec{n})=\mu(\hat{E}_M|x,n_1,\dots,n_M),\end{aligned} \tag{51}$$

and the conditional probability of the first k indices $n_1,\dots,n_k$ depends only on the first k projectors, that is,

$$\begin{gathered}P_c(n_1,\dots,n_k|x,\mathcal{M},\tau)=\\ P_c(n_1,\dots,n_k|x,\hat{E}_1,\dots,\hat{E}_k,\tau),\quad k\in\{1,\dots,M\}.\end{gathered} \tag{52}$$

We conclude this section by discussing a relation between the hypothesis of finiteness of the contextual information and a long-standing debate about the nature of the quantum state, which reached its apex with the Pusey-Barrett-Rudolph theorem [13]. In the framework of onthological theories, we can distinguish two possible cases. In one case, the quantum state is part of the classical description, so that the quantum state can be inferred by knowing the ontic state x. More precisely, two distributions $\rho(x|\psi,\eta)$ and $\rho(x|\psi',\eta')$ with $\psi\ne\psi'$ are not overlapped. In the other case, the inference of the quantum state from the ontic state is not generally possible. In Ref. [13], it was proved that the second case takes to a contradiction under a hypothesis of separability. Namely, the PBR hypothesis states, shortly speaking, that two spatially separate systems prepared in two quantum states (so that the overall quantum state is the product of the states) is associated with statistically independent classical variables. If the quantum state is taken as part of the classical description, then there are scenarios involving multiple measurements such that the information about the context is infinite. Thus, our hypothesis of finiteness of information has to lead to a break of the PBR separability condition under some scenario involving distinct systems and multiple measurements.

V. CONCLUSION AND PERSPECTIVES

In this paper, we have presented a generalization of the Gleason theorem and illustrated its application by deriving some general properties of a special class of HV theories. Apart from their relevance in quantum foundations, these theories are important also in quantum communication complexity as classical simulation protocols of quantum channels [11, 12]. Assuming that the amount of relevant information about the measurement context is finite, we have proved that the probability of an event for a single measurement with more than two outcomes must be linear in the projector associated with the event, given the information about the context. Further properties can be deduced considering multiple commuting measurements under the assumption of causality. This generalization of the Gleason theorem can suggest some clues for finding a classical model that replaces the quantum communication of n qubits with a finite amount of classical communication. At the present, this model is missing, apart from the Toner-Bacon model for single qubit [11] and a two-way communication model reported in Ref. [12]. We found the lower bound $2^n - 1$ for the amount of classical one-way communication required by an exact simulation [14] of n qubits. We have also discussed a possible relation of this work with the long-lasting debate on the nature of the quantum state (see Ref. [13] and references in there).

We conclude by suggesting some other possible extensions of this work. The proof of the generalized Gleason theorem requires that the measure μ is a generalized function [9], for which the derivatives are well-defined. This condition replaces the non-negativity condition used by Gleason, which has the same effect of ruling out highly discontinuous functions. Although our hypothesis is physically reasonable, it make the generalized Gleason theorem partially complementary to the original theorem. It would be interesting to find an alternative proof that requires only the non-negativity hypothesis and possibly uses an even weaker hypothesis on the set O (see theorem 3). It is worth noting that the concept of contextuality also applies to the preparation procedure [6], thus we could wonder if also this kind of contextuality can be summarized by a finite amount of information. In such a case, we would find that, given this partial information on the preparation context, the probability distribution of the ontic state x should be quadratic in the quantum state, like a quasi-probability distribution (such as Glauber-Sudarshan P distribution). In other terms, the probability distribution $\rho(x|\psi,\eta)$ would be somehow piecewise quadratic. Indeed, in Ref. [15] we proved that $\rho(x|\psi,\eta)$ cannot be quadratic on the whole Hilbert space, which is equivalent to say that a HV theory is contextual for a state preparation procedure, as remarked in Ref. [16].

VI. ACKNOWLEDGMENTS

We thank Xavier Coiteux-Roy for a careful reading of the manuscript and helpful comments. This work was supported by Swiss National Science Foundation (SNF).

[1] J. S. Bell, Physics **1**, 195 (1964).
[2] J. S. Bell, “Speakable and Unspeakable in Quantum Mechanics: Collected Papers”, (Cambridge University Press, 2004).
[3] S. Kochen and E. Specker, J. Math. Mech. **17**, 59 (1967).
[4] J. Barrett, L. Hardy, A. Kent, Phy. Rev. Lett. **95**, 010503 (2005).
[5] N. D. Mermin, Rev. Mod. Phys. **65**, 803 (1993).
[6] R. W. Spekkens, Phys. Rev. A **71**, 052108 (2005).
[7] Z. Chen, A. Montina, Phys. Rev. A **83**, 042110 (2011).
[8] A. M. Gleason, J. Math. Mech. **6**, 885-893 (1957); A. Peres, *Quantum Theory: Concepts and Methods* (Kluwer, Dordrecht, 1995).
[9] A. Friedman, *Generalized Functions and Partial Differential Equations* (Dover Publications, Inc., Mineola, New York, 2005).
[10] A. Montina, Phys. Rev. A **84**, 042307 (2011).
[11] B. F. Toner, D. Bacon, Phys. Rev. Lett. **91**, 187904 (2003).
[12] S. Massar, D. Bacon, N. J. Cerf, R. Cleve, Phys. Rev. A **63**, 052305 (2001).
[13] M. F. Pusey, J. Barrett, T. Rudolph, Nature Phys. **8** (6), 475.
[14] A. Montina, Phys. Rev. A **84**, 060303(R) (2011).
[15] A. Montina, Phys. Rev. Lett. **97**, 180401 (2006).
[16] R. W. Spekkens, Phys. Rev. Lett. **101**, 020401 (2008).