Contextuality for preparations, transformations, and unsharp measurements

R. W. Spekkens*

Perimeter Institute for Theoretical Physics, 35 King St. North, Waterloo, Ontario N2J 2W9, Canada
(Received 6 September 2004; published 31 May 2005)

The Bell-Kochen-Specker theorem establishes the impossibility of a noncontextual hidden variable model of quantum theory, or equivalently, that quantum theory is contextual. In this paper, an operational definition of contextuality is introduced which generalizes the standard notion in three ways: (i) it applies to arbitrary operational theories rather than just quantum theory, (ii) it applies to arbitrary experimental procedures rather than just sharp measurements, and (iii) it applies to a broad class of ontological models of quantum theory rather than just deterministic hidden variable models. We derive three no-go theorems for ontological models, each based on an assumption of noncontextuality for a different sort of experimental procedure; one for preparation procedures, another for unsharp measurement procedures (that is, measurement procedures associated with positive-operator valued measures), and a third for transformation procedures. All three proofs apply to two-dimensional Hilbert spaces, and are therefore stronger than traditional proofs of contextuality.

I. INTRODUCTION

Traditionally, a *noncontextual* hidden variable model of quantum theory is one wherein the measurement outcome that occurs for a particular set of values of the hidden variables depends only on the Hermitian operator associated with the measurement and not on which Hermitian operators are measured simultaneously with it. For instance, suppose A, B, and C are Hermitian operators such that A and B commute, A and C commute, but B and C do not commute. Then the assumption of noncontextuality is that the value predicted to occur in a measurement of A does not depend on whether B or C was measured simultaneously. The Bell-Kochen-Specker theorem shows that a hidden variable model of quantum theory that is noncontextual in this sense is impossible for Hilbert spaces of dimension 3 or greater [1–3].

The traditional definition of noncontextuality is lacking in several respects: (i) it does not apply to an arbitrary physical theory, but is rather specific to quantum theory; (ii) it does not apply to unsharp measurements, that is, those associated with positive-operator valued measures (POVMs), nor does it apply to preparation or transformation procedures; and (iii) it does not apply to ontological models wherein the outcomes of measurements are determined only probabilistically from the complete physical state of the system under investigation, for instance, indeterministic hidden variable models or ontological models of quantum theory lacking hidden variables. In this paper, we propose a new definition:

A noncontextual ontological model of an operational theory is one wherein if two experimental procedures are operationally equivalent, then they have equivalent representations in the ontological model.

This definition will be explained in Sec. II, where we provide a precise account of what it is for two experimental procedures to be operationally equivalent, and describe what is meant by an ontological model of an operational theory, specifying in particular how different experimental proce-

dures (preparations, measurements, and transformations) are represented in such a model. We also explain why it is appropriate to call this sort of ontological model *noncontextual* by providing an operational definition of an experimental context.

In Sec. III, we specialize our definition to the case of quantum theory. We provide examples of the sorts of contexts that can arise for preparations, transformations, and measurements, and describe what an assumption of noncontextuality for each type of procedure implies for an ontological model of quantum theory. In the case of measurements, we also generalize the object that is examined for context dependence from outcomes to probabilities of outcomes, and discuss the motivation for doing so. Further, we show how the traditional notion of noncontextuality is subsumed as a special case of our generalized notion when the outcomes of sharp measurements are assumed to be uniquely determined by the complete physical state of the system under investigation.

In Secs. IV, V, and VI, we provide no-go theorems for ontological models based on the assumption of noncontextuality for preparations, unsharp measurements, and transformations, respectively. All three proofs apply to two-dimensional (2D) Hilbert spaces, and are therefore stronger than traditional no-go theorems for noncontextuality, which require Hilbert spaces of dimension 3 or greater.[1] In Sec. V, we also provide a no-go theorem for noncontextuality of unsharp measurements based on a recent generalization of Gleason's theorem to 2D Hilbert spaces [5,6]. Section VII provides a general discussion of the motivation and plausibility of noncontextuality for different sorts of procedures, and Sec. VIII investigates the connection between these dif-

---

[1]Recent works by Cabello [4] and Busch [5] generalize the notion of contextuality to unsharp measurements in a manner that is different from the proposal of this paper. From our perspective, this work makes use of an assumption of deterministic outcomes for unsharp measurements that cannot be justified by an assumption of noncontextuality. This issue is discussed in Sec. VIII.

*Electronic address: rspekkens@perimeterinstitute.ca

ferent sorts of noncontextuality and the assumption that measurement outcomes are uniquely determined by the complete physical state of the system under investigation. Some conclusions and questions for future research are presented in Sec. IX.

II. DEFINITIONS OF NONCONTEXTUALITY FOR ANY OPERATIONAL THEORY

In an operational interpretation of a physical theory, the primitive elements are preparation procedures, transformation procedures, and measurement procedures. These are understood as lists of instructions to be implemented in the laboratory. The role of an operational theory is merely to specify the probabilities $p(k|P,T,M)$ of different outcomes k that may result from a measurement procedure M given a particular preparation procedure P, and a particular transformation procedure T. When there is no transformation procedure, or when it is considered to be part of the preparation or the measurement, we have simply $p(k|P,M)$.

Given the rule for determining probabilities of outcomes, one can define a notion of equivalence among experimental procedures. Specifically, two preparation procedures are deemed equivalent if they yield the same long-run statistics for every possible measurement procedure, that is, P is equivalent to P' if

$$p(k|P,M) = p(k|P',M) \text{ for all } M. \tag{1}$$

Two measurement procedures are deemed equivalent if they yield the same long-run statistics for every possible preparation procedure, that is, M is equivalent to M' if their outcomes can be associated one-to-one such that

$$p(k|P,M) = P(k|P,M') \text{ for all } P. \tag{2}$$

Finally, two transformation procedures are deemed equivalent if they yield the same long-run statistics for every possible preparation procedure that may precede and every possible measurement procedure that may follow, that is, T is equivalent to T' if

$$p(k|P,T,M) = p(k|P,T',M) \text{ for all } P,M. \tag{3}$$

It follows that one can distinguish two types of features of an experimental procedure: the first type of feature is one that is specified by specifying the equivalence class that the procedure falls in, while the second type is one that is not. The set of features of the second type—those that are not specified by specifying the equivalence class—we call the *context* of the experimental procedure. Note that, by our definition of an experimental context, having knowledge of the context does not enable one to predict the outcome of an experiment any better than if one only knew the equivalence class of the experimental procedure.

An example from quantum theory should clarify the notion of a context. Consider the following different measurement procedures for photon polarization. The first, which we denote by M_1, constitutes a piece of polaroid oriented to pass light that is vertically polarized along the $\hat{z}$ axis, followed by a photodetector. The second, which we denote by M_2, con-

stitutes a birefringent crystal oriented to separate light that is vertically polarized along the $\hat{z}$ axis from light that is horizontally polarized along this axis, followed by a photodetector in the vertically polarized output. The third and fourth procedures, denoted M_3 and M_4, are identical to M_1 and M_2, respectively, except that they are defined relative to an axis $\hat{n}$ that is skew to the $\hat{z}$ axis. It turns out that the statistics of outcomes for M_1 are the same as those for M_2, for all preparation procedures, and those for M_3 are the same as those for M_4. However, the statistics of outcomes for the first pair are different from those of the second. Thus, M_1 and M_2 fall in one equivalence class of measurements, and M_3 and M_4 fall in another. The orientation of the polaroid or calcite crystal is an example of the first sort of feature of an experimental operation, one whose variation involves a variation in the operational equivalence class of the procedure. On the other hand, whether one uses a piece of polaroid or a birefringent crystal to measure photon polarization is a feature of the measurement procedure of the second type; a variation of this feature does not change the equivalence class of the procedure. It is therefore part of the context of the measurement procedure.

To properly define a noncontextual ontological model of an operational theory, it is not enough to have a definition of context; we also need to specify precisely what we mean by an ontological model. We turn to this now.

An ontological model is an attempt to offer an explanation of the success of an operational theory by assuming that there exist physical systems that are the subject of the experiment. These systems are presumed to have attributes regardless of whether they are being subjected to experimental test, and regardless of what anyone knows about them. These attributes describe the real state of affairs of the system. Thus a specification of which instance of each attribute applies at a given time we call the *ontic state* of the system. If the ontic state is not completely specified after specifying the preparation procedure, then the additional variables required to specify it are called *hidden variables*. Although most ontological models do involve hidden variables, this is not always the case. For instance, the ontic states may be associated one-to-one with the equivalence classes of preparation procedures (as is the case for pure preparation procedures in the Beltrametti-Bugajski model of quantum theory [7]). We shall denote the complete set of variables in an ontological model by λ, and the space of values of λ by Ω.

Within an ontological model of an operational theory, preparation procedures are preparations *of* the ontic state of the system. However, the procedure need not fix this state uniquely; rather, it might only fix the probabilities that the system be in different ontic states. Thus someone who knows that a system was prepared using the preparation procedure P describes the system by a probability density $\mu_P(\lambda)$ over the model variables.

Similarly, measurement procedures are measurements *of* the ontic state of the system. Again, these procedures need not enable one to infer the identity of the ontic state uniquely, nor need they even enable one to infer a set of ontic states within which the actual ontic state lies. Rather, they might only enable one to infer probabilities for the system to have been in different ontic states. In this, the most general case,

the outcome of the measurement is not uniquely determined by λ. Only the probabilities of the different outcomes are so determined. Thus, for every value of λ, one associates a probability $\xi_{M,k}(\lambda)$ which is the probability of obtaining outcome k in a measurement M given that the system is in the ontic state λ. We call $\xi_{M,k}(\lambda)$ an "indicator function."[2]

Finally, transformation procedures are transformations *of* the ontic state of the system. These may be stochastic transitions. Thus a transformation procedure T is represented by a transition matrix, $\Gamma_T(\lambda',\lambda)$, which represents the probability density for a transition from the ontic state λ to the ontic state λ'.

Thus within an ontological model, if the preparation procedure is P, and the measurement procedure is M, then the probability assigned to outcome k is the probability assigned to outcome k given λ, averaged over all λ, weighted by the probability of λ, that is, $\int d\lambda \mu_P(\lambda)\xi_{M,k}(\lambda)$. If there is a transformation procedure T intervening between the preparation and measurement, and this is associated with the transition matrix $\Gamma_T(\lambda',\lambda)$, then the probability of outcome k is $\int d\lambda' d\lambda \xi_{M,k}(\lambda')\Gamma_T(\lambda',\lambda)\mu_P(\lambda)$.

Summarizing, an ontological model assumes: (i) every preparation procedure P is associated with a normalized probability density over the ontic state space, $\mu_P:\Omega \rightarrow [0,1]$ such that $\int \mu_P(\lambda)d\lambda = 1$; (ii) every measurement procedure M with outcomes labeled by k is associated with a set of indicator functions $\{\xi_{M,k}(\lambda)\}_k$ over the ontic states, that is, a set of functions $\xi_{M,k}:\Omega \rightarrow [0,1]$ satisfying $\Sigma_k \xi_{M,k}(\lambda)=1$ for all λ; (iii) every transformation procedure T is associated with a transition matrix $\Gamma_T:\Omega \times \Omega \rightarrow [0,1]$ such that $\int \Gamma_T(\lambda',\lambda)d\lambda'=1$ for all λ, and (iv) the predictions of the operational theory are reproduced exactly by the model, that is,

$$p(k|P,T,M) = \int d\lambda' d\lambda \xi_{M,k}(\lambda')\Gamma_T(\lambda',\lambda)\mu_P(\lambda) \qquad (4)$$

for all P, T, and M.

In general, the representation of an experimental procedure in an ontological model might depend on both its equivalence class and its context.[3] It is natural, however, to consider the possibility of a model wherein the representa-

tion of every experimental procedure depends *only* on its equivalence class and not on its context. After all, a natural way to explain the fact that a pair of preparation (measurement, transformation) procedures are operationally equivalent is to assume that they prepare (measure, transform) the ontic state of the system in precisely the same way. We shall call such an ontological model *noncontextual*. Any operational theory that admits such a model shall also be called *noncontextual*.[4] In general, if any set of procedures is represented in a context-independent way within an ontological model, we shall say that the model is *noncontextual* for those procedures.

It is useful to explicitly characterize the assumption of noncontextuality for preparations, transformations, and measurements. We will call an ontological model *preparation noncontextual* if the representation of every preparation procedure is independent of context, that is, if

$$\mu_P(\lambda) = \mu_{e(P)}(\lambda), \qquad (5)$$

where $e(P)$ is P's equivalence class. Similarly, we will call a model *measurement noncontextual* if the representation of every measurement procedure is independent of context, that is, if

$$\xi_{M,k}(\lambda) = \xi_{e(M),k}(\lambda), \qquad (6)$$

where $e(M)$ is M's equivalence class. Finally, a model is called *transformation noncontextual* if the representation of every transformation procedure is independent of context, that is, if

$$\Gamma_T(\lambda',\lambda) = \Gamma_{e(T)}(\lambda',\lambda), \qquad (7)$$

where $e(T)$ is T's equivalence class. A *universally noncontextual* ontological model is one that is noncontextual for all experimental procedures: preparations, transformations, and measurements.

III. DEFINITIONS OF NONCONTEXTUALITY IN QUANTUM THEORY

We begin with a quick review of the operational approach to quantum theory, described, for instance, in Refs. [9–12].

An equivalence class of preparation procedures is associated with a density operator ρ. This is a positive trace-1 operator over the Hilbert space $\mathcal{H}$ of the system: $\rho > 0$, $\mathrm{Tr}(\rho)=1$. Rank-1 density operators are simply projectors onto rays of Hilbert space, and are called *pure*.

An equivalence class of measurement procedures is associated with a positive operator valued measure (POVM) $\{E_k\}$. A POVM is an ordered set $\{E_k\}$ of positive operators that sum to identity, $\Sigma_k E_k=I$. The kth element E_k is associated with the kth outcome. Specifically, given a preparation asso-

---

[2]Some might argue that within the framework of an ontological model, the term *measurement* ought to be reserved for a procedure which reveals some attribute of the system under investigation. However, we feel that it is suitable for any procedure that leads to an update in one's information about which instance of some attribute applied, or equivalently, in one's information about what the ontic state of the system was prior to the procedure being implemented. Note that even this weak notion of what constitutes a measurement fails for some experimental procedures that lead to distinct outcomes, namely, those wherein the probabilities of the different outcomes are independent of the ontic state. For simplicity, however, we shall not introduce any novel terminology for this exceptional case.

[3]If one allows such generality, then—perhaps contrary to a common impression—it is possible to provide an ontological model of quantum theory. The deBroglie-Bohm theory [8] is an example.

[4]This terminology allows one to use the phrase "quantum theory is contextual" as a shorthand for "quantum theory does not admit a noncontextual ontological model," much as it is common to use the phrase "quantum theory is nonlocal" in place of "quantum theory does not admit a local ontological model."

ciated with a density operator ρ, the probability of the outcome k is simply $\text{Tr}(\rho E_k)$. It is useful to single out the POVMs whose elements are idempotent, that is, those for which $E_k^2 = E_k$ for all k. Since idempotent positive operators are projectors, these POVMs are called *projective-valued* measures (PVMs). The associated measurements are said to be *sharp*. These are the sorts of measurements that are considered in standard textbook treatments of quantum mechanics. A Hermitian operator defines a PVM through the projectors in its spectral resolution.

Finally, an equivalence class of transformation procedures is associated with a completely positive (CP) map $\mathcal{T}$. A CP map $\mathcal{T}$ is a positive linear map on the space of operators over $\mathcal{H}$ such that $\mathcal{T} \otimes \mathcal{I}$ is a positive linear map on the space of operators over $\mathcal{H} \otimes \mathcal{H}'$, where $\mathcal{H}'$ is of arbitrary dimension, and $\mathcal{I}$ is the identity map on $\mathcal{H}'$. Unitary maps, familiar from textbook treatments of quantum theory, are reversible CP maps.

A preparation procedure associated with a non-rank-1 density operator ρ can be implemented in as many ways as there are convex decompositions of ρ. Suppose $\{(p_k, \rho_k)\}$ is a convex decomposition of ρ, that is, $\rho = \Sigma_k p_k \rho_k$. If one generates a random number according to the distribution p_k, and upon obtaining the number k, one implements the preparation associated with ρ_k, this procedure is a member of the equivalence class of procedures associated with ρ. Another way of implementing a preparation procedure that is associated with ρ is to implement a preparation of a purification $|\psi\rangle$ of ρ on $\mathcal{H} \otimes \mathcal{H}'$ (a purification of ρ is any state $|\psi\rangle$ such that $\text{Tr}_{\mathcal{H}'} |\psi\rangle\langle\psi| = \rho$). The equivalence class therefore also contains members associated with different purifications of ρ.

The assumption of preparation noncontextuality in quantum theory is that the probability distribution over ontic states that is associated with a preparation procedure P depends only on the density operator ρ associated with P,

$$\mu_P(\lambda) = \mu_\rho(\lambda). \qquad (8)$$

In particular, the distribution does not depend on the particular convex decomposition of ρ or on the particular purification of ρ that is used in the preparation procedure.

The multiplicity of contexts for transformation procedures parallels the multiplicity of contexts for preparation procedures. Transformation procedures that are associated with nonunitary CP maps can be obtained as a convex sum of unitary maps in many different ways, each one of which corresponds to a distinct transformation procedure, and can also be obtained by implementing a unitary map on a larger system that incorporates the system of interest [13].

The assumption of transformation noncontextuality in quantum theory is that the transition matrix that is associated with a given transformation procedure T depends only on the CP map $\mathcal{T}$ associated with T,

$$\Gamma_T(\lambda', \lambda) = \Gamma_\mathcal{T}(\lambda', \lambda). \qquad (9)$$

It does not, for instance, depend on the particular convex sum of unitaries or the particular unitary on a larger system by which the transformation was implemented.

In the case of quantum measurements, there are also many sorts of contexts. For instance, every fine-graining of a non-maximally informative measurement (i.e., a measurement associated with a POVM at least one element of which is not rank 1) provides a different context. Suppose the POVM $\{F_j\}$ is a fine-graining of $\{E_k\}$, which is to say that there is a partitioning of the outcomes j into sets S_k such that $E_k = \Sigma_{j \in S_k} F_j$. By implementing a measurement associated with the POVM $\{F_j\}$, then discarding all information about j except the set S_k to which it belongs, one implements a measurement in the equivalence class associated with the POVM $\{E_k\}$.

Despite the fact that the independence of representation on fine-graining is traditionally the full extent of the assumption of noncontextuality for measurements (as we will show below), it is not difficult to see that there are many other sorts of contexts. For instance, there is a context for every convex decomposition of a nonmaximal measurement. A convex decomposition of a POVM $\{E_k\}$ is defined as a probability distribution $\{p_\alpha\}$ and a set of POVMs, $\{\mathcal{F}_\alpha\}$ where $\mathcal{F}_\alpha = \{F_k^\alpha\}$, such that $E_k = \Sigma_\alpha p_\alpha F_k^\alpha$. By sampling α from the distribution $\{p_\alpha\}$, then implementing the measurement associated with the POVM $\{F_k^\alpha\}$, and registering only the outcome of this measurement, one implements a measurement in the equivalence class associated with the POVM $\{E_k\}$. There is also a context for every way of obtaining a POVM by coupling to an ancilla and measuring a PVM on the composite of system+ancilla [14].

The assumption of measurement noncontextuality in quantum theory is that the set of indicator functions representing a measurement M depends only on the POVM $\{E_k\}$ associated with M,

$$\xi_{M,k}(\lambda) = \xi_{\{E_k\},k}(\lambda). \qquad (10)$$

In addition to admitting new sorts of contexts for measurements, our generalized notion of measurement contextuality involves a slight revision of *what it is* that depends on the measurement context.

In the past, measurement contextuality has only been considered within the framework of deterministic hidden variable theories, and the question of interest has been whether or not the measurement outcome for a given ontic state of the system depends on the context of the measurement. However, for *objectively indeterministic* ontological models, it is clear that the natural question to ask is whether the *probabilities* of different outcomes for a given ontic state of the system depend on the context. This is analogous to Bell's [15] generalization of the notion of locality from measurement outcomes being causally independent of parameter settings at spacelike separation to the *probabilities* of measurement outcomes being causally independent of parameter settings at spacelike separation.[5] This distinction was introduced by Bell in order to cleanly separate the notion of locality from

---

[5] More specifically, Bell [15] defined a theory to be *locally deterministic* if the variables in space-time region I are determined by the variables in a space-time region that fully closes the backward light cone of I, and *locally causal* if the probability distribution over

the notion of determinism. Similarly, our generalized definition allows one to cleanly separate the notion of measurement noncontextuality from the notion of determinism.

Once the question is posed, it is somewhat obvious that if there is to be any notion of measurement contextuality within objectively indeterministic ontological models, the appropriate quantities to examine for context dependence are the *probabilities* of different outcomes for a given ontic state of the system. A less obvious feature of our generalized notion of measurement contextuality is that the probabilities of outcomes are the appropriate quantities to examine for context dependence even in *objectively deterministic* ontological models. The key is that the latter sort of model may still exhibit an *epistemic* indeterminism, wherein knowledge of the equivalence class of the measurement together with the ontic state of the system under investigation does not uniquely fix the outcome. To explain this properly, we need to consider some of the details of the mathematical representation of these measurements.

The distinction between the ontic state of the system determining the outcome and determining only the probabilities of different outcomes is captured mathematically within an ontological model by the sorts of indicator functions one uses to represent measurements. The former case is represented by an indicator function that is *idempotent*, that is, one for which $\chi(\lambda)^2 = \chi(\lambda)$ [we shall denote idempotent indicator functions by $\chi(\lambda)$ rather than $\xi(\lambda)$]. Such functions are necessarily equal to 1 in some region of the ontic state space and 0 elsewhere. By virtue of the fact that a set of indicator functions must satisfy $\Sigma_k \chi_k(\lambda) = 1$, if all the indicator functions are idempotent, then they must be nonoverlapping, that is, $\chi_k(\lambda)\chi_{k'}(\lambda) = 0$ for $k \neq k'$. Thus for every value of λ, only a single indicator function in the set $\{\chi_k(\lambda)\}$ receives the value 1 while the others receive the value 0. Since the value of the kth indicator function at a given λ specifies the probability of the kth outcome given the ontic state λ, the outcome of the measurement is determined for all ontic states if and only if the latter is represented by a set of idempotent indicator functions.

We shall call the assumption that a particular measurement is represented by a set of idempotent indicator functions the assumption of *outcome determinism* for that measurement.

Now note that even within an objectively deterministic ontological model, measurements may fail to exhibit outcome determinism: specifying the ontic state of the system under investigation together with the equivalence class of the measurement procedure may be insufficient to uniquely fix the outcome. The outcome might only be fixed uniquely by supplementary features of the measurement procedure (which constitute part of the context of the measurement by our definition), such as microscopic degrees of freedom of the apparatus. Because the indicator function for a measure-

ment specifies the dependence of the outcome on the ontic state of the system under investigation, and not the dependence of the outcome on the ontic state of any systems that make up the measurement apparatus or the environment, such a measurement must be represented by a nonidempotent set of indicator functions. Nonetheless, it may still be the case that for each equivalence class of measurements, all the elements of the class are represented by the *same* nonidempotent set of indicator functions, and *this* is all that is required for the measurements to be deemed noncontextual by our definition.

As an example, consider a classical system and a classical measurement device that generates an outcome by rolling one of several differently weighted dice, with the choice of the dice being determined by the ontic state of the system. Two such devices are only found to be operationally equivalent if all of the dice of one are weighted in the same way as those of the other. Thus every device in the equivalence class is represented by the same set of indicator functions, and consequently one has measurement noncontextuality by our definition. The underlying ontological model (classical mechanics) is objectively deterministic, but in order to predict the outcome of a particular measurement, one must supplement the ontic state of the system by the precise initial configuration of the dice and their environment, features that form part of the context of the measurement. Thus although the outcome of the measurement clearly depends on the context, we take this to be a failure of outcome determinism rather than a failure of measurement noncontextuality.[6]

Thus it is really the notion of *outcome determinism*, rather than the notion of determinism, which we seek to cleanly separate from the notion of measurement noncontextuality through our generalized definition. This makes our definition of measurement noncontextuality revisionist insofar as the traditional definition implicitly incorporated the assumption of outcome determinism, while ours does not. This suggested revision in terminology is motivated by the idea that what is crucial to the notion of a noncontextual ontological model is that it reproduces the equivalence class structure of the operational theory.

It is worth noting that, given the additional assumption of outcome determinism, one can recover the traditional definition of measurement noncontextuality as a special case of our definition. Specifically, if one considers only sharp measurements and one represents these by sets of idempotent functions (i.e., one assumes outcome determinism for these measurements), then the assumption of the independence of the representation of a measurement on the fine-graining of the PVM with which it is implemented is just the traditional notion of noncontextuality (described in the Introduction). This can be seen as follows. Specifying whether a Hermitian operator A is measured together with B or with C is equivalent to specifying a fine-graining of the PVM $\{\Pi_k\}$ that is

---

values for a variable in space-time region I are determined by a specification of the values of all the variables in the backward light cone of I ("determined" in the sense that further conditioning on variables in the region outside the backward light cone would not change the probability distribution).

[6]By our definition, any classical theory is necessarily noncontextual for all experimental procedures. This highlights another virtue of our particular definition: contextuality, in all of its manifestations, is found to be a nonclassical phenomenon. For an opposing perspective, see Ref. [16].

defined by the spectral resolution of A; the simultaneous eigenspaces of A and B define one such fine-graining, while the simultaneous eigenspaces of A and C define another. Specifying the eigenvalue assigned to an operator A for every value of λ is equivalent to specifying a set of idempotent indicator functions; the values of λ in the support of the function associated with outcome k are simply those that assign the kth eigenvalue to A. Clearly then, assuming that the value assigned to A is independent of whether A is measured together with B or C is equivalent to assuming that the set of idempotent indicator functions associated with the PVM $\{\Pi_k\}$ is independent of the fine-graining by which it was implemented.

No-go theorems based on the traditional definition of noncontextuality apply only in Hilbert spaces of dimensionality 3 or greater. Moreover, one cannot extend such proofs to 2D Hilbert spaces because there are no fine-grainings of nontrivial PVM measurements in a 2D Hilbert space, and fine-graining is the only notion of context that is recognized traditionally. However, by appealing to preparations, transformations, and unsharp measurements, which admit many contexts even in a 2D Hilbert space, proofs of contextuality can be achieved here as well. From this perspective, the restriction of previous proofs of contextuality to 3D Hilbert spaces was an artifact of a limited notion of a context.

Among the new proofs of contextuality that we shall present, the proof of preparation contextuality is the simplest, and so we begin with this case.

IV. PROOF OF PREPARATION CONTEXTUALITY IN TWO DIMENSIONS

There are two features of the representation of preparation procedures in an ontological model that are central to our proof. The first concerns distinguishability, and the second convex combination.

Feature 1. If two preparation procedures, P and P' are distinguishable with certainty in a single-shot measurement, then their associated probability distributions, $\mu(\lambda)$ and $\mu'(\lambda)$, are nonoverlapping, that is,

$$\mu(\lambda)\mu'(\lambda) = 0 \text{ for all } \lambda. \tag{11}$$

This feature can be understood as follows. Suppose one wishes to perform a measurement that discriminates, with certainty, between two probability distributions. In other words, one wishes to perform a measurement that allows one to retrodict, with certainty, which distribution applied. This is only possible if the distributions to be discriminated are nonoverlapping. The reason is that if the two distributions overlapped in some region of the space of ontic states, then whenever the actual ontic state was in that region (and it will sometimes be in that region, because the region is assigned nonzero probability by both distributions), no measurement would be able to distinguish with certainty whether the system had been prepared using one or the other distribution, since the actual ontic state is consistent with both. Thus if, within an operational theory, a pair of preparation procedures are distinguishable with certainty, then the only way an ontological model of the theory can account for this fact is by

associating these procedures with nonoverlapping distribution functions.

The second feature of an ontological model that is critical to our proof is the manner in which convex combinations of preparation procedures are represented. Suppose that the preparation procedures P and P' are represented by distributions $\mu(\lambda)$ and $\mu'(\lambda)$. Now suppose that a bit is generated uniformly at random from the distribution p, $1-p$, and the value of the bit is used to determine whether P or P' is implemented, after which the bit is forgotten. This effective procedure, which we call P'', must be represented within the ontological model by a distribution $\mu''(\lambda)$ satisfying

$$\mu''(\lambda) = p\mu(\lambda) + (1-p)\mu'(\lambda). \tag{12}$$

The reason is as follows. The probability that the ontic state of the system is λ given procedure P'', is simply the sum of the probability that it is λ given procedure P and the probability that it is λ given procedure P', weighted by the respective probabilities of P and P' given P''.

Thus we have the following.

Feature 2. A convex combination of preparation procedures is represented within an ontological model by a convex sum of the associated probability distributions.

With these facts in hand, we now proceed with the proof.

Consider a set of six pure preparations, denoted P_a, P_A, P_b, P_B, P_c, and P_C, corresponding to the normalized Hilbert space vectors,

$$\psi_a = (1,0),$$

$$\psi_A = (0,1),$$

$$\psi_b = (1/2, \sqrt{3}/2),$$

$$\psi_B = (\sqrt{3}/2, -1/2),$$

$$\psi_c = (1/2, -\sqrt{3}/2),$$

$$\psi_C = (\sqrt{3}/2, 1/2), \tag{13}$$

or, equivalently, the rank 1 density operators,

$$\sigma_a = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

$$\sigma_A = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$$

$$\sigma_b = \begin{pmatrix} \dfrac{1}{4} & \dfrac{1}{4}\sqrt{3} \\ \dfrac{1}{4}\sqrt{3} & \dfrac{3}{4} \end{pmatrix},$$

$$\sigma_B = \begin{pmatrix} \dfrac{3}{4} & -\dfrac{1}{4}\sqrt{3} \\ -\dfrac{1}{4}\sqrt{3} & \dfrac{1}{4} \end{pmatrix},$$

$$\sigma_c = \begin{pmatrix} \frac{1}{4} & -\frac{1}{4}\sqrt{3} \\ -\frac{1}{4}\sqrt{3} & \frac{3}{4} \end{pmatrix},$$

$$\sigma_C = \begin{pmatrix} \frac{3}{4} & \frac{1}{4}\sqrt{3} \\ \frac{1}{4}\sqrt{3} & \frac{1}{4} \end{pmatrix}. \tag{14}$$

One can easily verify the following orthogonality conditions:

$$\sigma_a \sigma_A = 0, \tag{15}$$

$$\sigma_b \sigma_B = 0, \tag{16}$$

$$\sigma_c \sigma_C = 0. \tag{17}$$

Now consider the preparation procedure wherein one of P_a or P_A is implemented, with the choice being made uniformly at random (for instance, by flipping a fair coin), and with no record being made of the choice. Denote this procedure by P_{aA}. Define procedures P_{bB} and P_{cC} similarly. Consider also the preparation procedure wherein one of P_a, P_b, or P_c is implemented, again, with equal probabilities for each, and without recording the choice. We denote this by P_{abc}. The procedure P_{ABC} is defined similarly.

These procedures are represented in the quantum formalism by the appropriate convex sums of the density operators in Eq. (14). It turns out that all of these convex sums yield the same rank-2 density operator, namely,

$$I/2 = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}, \tag{18}$$

commonly referred to as the "completely mixed state." Specifically,

$$I/2 = \frac{1}{2}\sigma_a + \frac{1}{2}\sigma_A \tag{19}$$

$$= \frac{1}{2}\sigma_b + \frac{1}{2}\sigma_B \tag{20}$$

$$= \frac{1}{2}\sigma_c + \frac{1}{2}\sigma_C \tag{21}$$

$$= \frac{1}{3}\sigma_a + \frac{1}{3}\sigma_b + \frac{1}{3}\sigma_c \tag{22}$$

$$= \frac{1}{3}\sigma_A + \frac{1}{3}\sigma_B + \frac{1}{3}\sigma_C. \tag{23}$$

In Fig. 1, we present the Bloch ball representation of the seven density operators defined above. This provides a
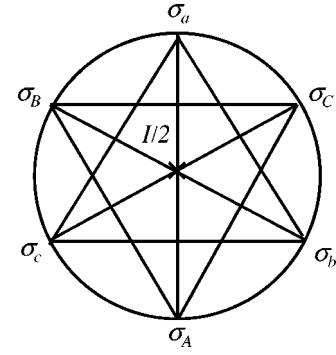

FIG. 1. The Bloch ball representation of the six pure states and the five convex decompositions of the completely mixed state used in the proof of preparation contextuality. Each convex decomposition is represented by a convex polytope whose vertices represent the elements of the decomposition [17]. The two-element decompositions in our example are represented by line segments, and the three-element decompositions by equilateral triangles.

graphical synopsis of the relevant orthogonality relations and convex structure.

Within an ontological model of operational quantum theory, each preparation procedure P_x is associated with a probability distribution $\mu_x(\lambda)$. Now note that if two density operators, σ and σ', are orthogonal in the vector space of operators, that is, $\sigma\sigma' = 0$, then the associated preparation procedures can be distinguished with certainty in a single-shot measurement (for instance, for preparations associated with orthogonal Hilbert space vectors, one simply implements the measurement associated with an orthogonal basis that includes these vectors). By feature 1 of ontological models (described above), distinguishable procedures are represented by nonoverlapping distributions. Thus from Eqs. (15)–(17) we can infer that

$$\mu_a(\lambda)\mu_A(\lambda) = 0, \tag{24}$$

$$\mu_b(\lambda)\mu_B(\lambda) = 0, \tag{25}$$

$$\mu_c(\lambda)\mu_C(\lambda) = 0. \tag{26}$$

Furthermore, in any ontological model a convex combination of preparation procedures is represented by a convex sum of the associated probability distributions (feature 2 above). Thus if the procedures $P_{aA}, P_{bB}, \dots, P_{ABC}$ are represented by distributions $\mu_{aA}(\lambda), \mu_{bB}(\lambda), \dots, \mu_{ABC}(\lambda)$, the manner in which these procedures are obtained by convex combination of $P_a, P_b, \dots, P_C$ implies that

$$\mu_{aA}(\lambda) = \frac{1}{2}\mu_a(\lambda) + \frac{1}{2}\mu_A(\lambda), \tag{27}$$

$$\mu_{bB}(\lambda) = \frac{1}{2}\mu_b(\lambda) + \frac{1}{2}\mu_B(\lambda), \tag{28}$$

$$\mu_{cC}(\lambda) = \frac{1}{2}\mu_c(\lambda) + \frac{1}{2}\mu_C(\lambda), \tag{29}$$

$$\mu_{abc}(\lambda) = \frac{1}{3}\mu_a(\lambda) + \frac{1}{3}\mu_b(\lambda) + \frac{1}{3}\mu_c(\lambda), \qquad (30)$$

$$\mu_{ABC}(\lambda) = \frac{1}{3}\mu_A(\lambda) + \frac{1}{3}\mu_B(\lambda) + \frac{1}{3}\mu_C(\lambda). \qquad (31)$$

The assumption of a preparation noncontextual ontological model is that the distribution associated with a preparation procedure depends only on the operational equivalence class of that procedure, and thus only on the density operator associated with that procedure. Since the procedures $P_{aA}, P_{bB}, \ldots, P_{ABC}$ are all represented by $I/2$, they must all be represented by the same distribution in a preparation noncontextual ontological model. Thus we require $\mu_{aA} = \mu_{bB} = \cdots = \mu_{ABC}$. Denoting this distribution by $\nu(\lambda)$, we have simply

$$\nu(\lambda) = \frac{1}{2}\mu_a(\lambda) + \frac{1}{2}\mu_A(\lambda) \qquad (32)$$

$$= \frac{1}{2}\mu_b(\lambda) + \frac{1}{2}\mu_B(\lambda) \qquad (33)$$

$$= \frac{1}{2}\mu_c(\lambda) + \frac{1}{2}\mu_C(\lambda) \qquad (34)$$

$$= \frac{1}{3}\mu_a(\lambda) + \frac{1}{3}\mu_b(\lambda) + \frac{1}{3}\mu_c(\lambda) \qquad (35)$$

$$= \frac{1}{3}\mu_A(\lambda) + \frac{1}{3}\mu_B(\lambda) + \frac{1}{3}\mu_C(\lambda). \qquad (36)$$

We now show that there is no set of distributions satisfying Eqs. (24)–(26) and Eqs. (32)–(36). Consider the values of the various probability densities at a fixed value of λ. We denote these simply as $\mu_a, \mu_A, \ldots, \mu_C$. We show that the only solution to all the constraints, for a fixed λ, is $\mu_a, \mu_A, \ldots, \mu_C = 0$, which we call the all-zero solution.

To satisfy Eqs. (24)–(26), one of the pair μ_a and μ_A must be zero, as must be one of the pair μ_b and μ_B and one of the pair μ_c and μ_C. In all, there are eight possible assignments of zeroes that satisfy Eqs. (24)–(26). We consider each of these in turn.

If we have $\mu_a, \mu_b, \mu_c = 0$ then by Eq. (35) we have $\nu = 0$, and by Eqs. (32)–(34), we conclude that $\mu_A, \mu_B, \mu_C = 0$, so that we have the all-zero solution. If, instead we have $\mu_a, \mu_b, \mu_C = 0$ then by combining Eq. (34) and Eq. (35) we find $\frac{1}{2}\mu_c = \frac{1}{3}\mu_c$, for which the only solution is $\mu_c = 0$. But this gets us back to the first case, and the all-zero solution. Every other case yields the all-zero solution by virtue of the symmetry of the problem under rotations by multiples of 60° in the Bloch sphere representation.

The above argument did not depend on λ, and thus for all λ the only solution is the all-zero solution. Consequently, the only set of distributions that satisfy Eqs. (24)–(26) and Eqs. (32)–(36) is the set of uniformly zero distributions, $\mu_a(\lambda), \mu_A(\lambda), \ldots, \mu_C(\lambda) = 0$. But such distributions are not probability distributions since they are not normalized to 1. This concludes the proof.

V. PROOFS OF CONTEXTUALITY FOR UNSHARP MEASUREMENTS IN TWO DIMENSIONS

Proofs of measurement contextuality have usually arisen only in the context of *sharp* measurements, that is, those associated with PVMs,[7] and outcome determinism has been assumed for such measurements. We shall make the same assumption here for sharp measurements, but we shall be considering unsharp measurements as well, that is, those associated with POVMs, and for these, outcome determinism will not be assumed. It is important to note that the "proofs of contextuality" presented in the next two subsections are contingent on the assumption of outcome determinism for sharp measurements. The status of this assumption will be revisited in Sec. VIII, where we will clarify what, precisely, has been proven.

A. A proof based on a finite set of measurements

Consider three binary-outcome measurements, M_a, M_b, and M_c, associated respectively with PVMs $\{\Pi_a, \Pi_A\}$, $\{\Pi_b, \Pi_B\}$, and $\{\Pi_c, \Pi_C\}$, where Π_a projects onto the ray spanned by ψ_a, Π_A projects onto the ray spanned by ψ_A, and so forth, with the vectors ψ_x being those that are defined in Eq. (13).[8]

By the definition of a PVM, we have

$$\Pi_a + \Pi_A = I, \qquad (37)$$

$$\Pi_b + \Pi_B = I, \qquad (38)$$

$$\Pi_c + \Pi_C = I, \qquad (39)$$

and

$$\Pi_a\Pi_A = 0, \qquad (40)$$

$$\Pi_b\Pi_B = 0, \qquad (41)$$

$$\Pi_c\Pi_C = 0. \qquad (42)$$

Given the assumption of outcome determinism for sharp measurements, the representations of M_a, M_b, and M_c in an ontological model are the sets of idempotent indicator functions $\{\chi_a(\lambda), \chi_A(\lambda)\}$, $\{\chi_b(\lambda), \chi_B(\lambda)\}$, and $\{\chi_c(\lambda), \chi_C(\lambda)\}$, respectively. By definition, these must satisfy

$$\chi_a(\lambda) + \chi_A(\lambda) = 1, \qquad (43)$$

$$\chi_b(\lambda) + \chi_B(\lambda) = 1, \qquad (44)$$

---

[7]Exceptions are Refs. [4,5,16,18].

[8]Note that $\Pi_a = \sigma_a$, $\Pi_A = \sigma_A$, etc., where σ_x is defined in Eq. (14). This follows from the fact that the rank-1 density operator associated with a vector is simply the projector onto the ray spanned by that vector. It follows that Eqs. (37)–(39) and Eqs. (40)–(42) are equivalent to Eqs. (19)–(21) and Eqs. (15)–(17), respectively. We use a distinct notation for the same mathematical operators to remind the reader of the fact that in this section they represent measurement outcomes rather than preparation procedures.

$$\chi_c(\lambda) + \chi_C(\lambda) = 1, \tag{45}$$

and

$$\chi_a(\lambda)\chi_A(\lambda) = 0, \tag{46}$$

$$\chi_b(\lambda)\chi_B(\lambda) = 0, \tag{47}$$

$$\chi_c(\lambda)\chi_C(\lambda) = 0. \tag{48}$$

Now consider choosing one of M_a, M_b, and M_c at random, with probability $1/3$ for each, implementing the chosen measurement, and only registering whether the first (small letter) or the second (capital letter) outcome occurred. Call the effective measurement procedure that results M. It is associated with the POVM

$$\left\{ \frac{1}{3}\Pi_a + \frac{1}{3}\Pi_b + \frac{1}{3}\Pi_c, \frac{1}{3}\Pi_A + \frac{1}{3}\Pi_B + \frac{1}{3}\Pi_C \right\}. \tag{49}$$

In an ontological model, a convex combination of measurements procedures is represented by an element-wise convex sum of the associated sets of indicator functions (for the same reason that an ontological model has feature 2 of Sec. IV). Thus M is represented by the set of indicator functions

$$\left\{ \frac{1}{3}\chi_a(\lambda) + \frac{1}{3}\chi_b(\lambda) + \frac{1}{3}\chi_c(\lambda), \frac{1}{3}\chi_A(\lambda) + \frac{1}{3}\chi_B(\lambda) + \frac{1}{3}\chi_C(\lambda) \right\}. \tag{50}$$

Note that the POVM (49) is equal to[9]

$$\left\{ \frac{1}{2}I, \frac{1}{2}I \right\}. \tag{51}$$

But it is clear from this way of writing the POVM that the measurement has a random outcome regardless of the preparation procedure, since $\mathrm{Tr}(\rho \frac{1}{2}I) = \frac{1}{2}$ regardless of ρ. It then follows that the equivalence class of measurement procedures that contains M also contains the "measurement" procedure $\tilde{M}$ that completely ignores the system and just flips a fair coin to determine the outcome. Now consider how the measurement $\tilde{M}$ is represented in the ontological model. Because the outcome does not depend on the system at all, it follows that regardless of the value of λ, there is a probability of $1/2$ for each outcome, so it is represented by the set of indicator functions

$$\left\{ \frac{1}{2}, \frac{1}{2} \right\}, \tag{52}$$

where each element should be thought of as a uniform function over λ of height $\frac{1}{2}$.[10]

By the assumption of measurement noncontextuality, the measurement M must be represented by the same set of in-

---

[9]This fact is also captured by Eqs. (22) and (23).

[10]This fact can also be established by noting that the equivalence class includes the measurement M', obtained from M by permuting the two outcomes (because such a permutation does not change the statistics of outcomes). The ontological representations of M and M' are $\{\xi_1(\lambda), \xi_2(\lambda)\}$ and $\{\xi_2(\lambda), \xi_1(\lambda)\}$. Now, the assumption of

dicator functions as the measurement $\tilde{M}$. It follows that the set of functions (50) must be equal to the set of functions (52). However, this constraint is inconsistent with the constraints (43)–(48). To satisfy Eqs. (43)–(48) it is necessary that for every value of λ, one of $\chi_a(\lambda)$ and $\chi_A(\lambda)$ must be equal to 0 and the other equal to 1. The same is true of $\chi_b(\lambda)$ and $\chi_B(\lambda)$ and of $\chi_c(\lambda)$ and $\chi_C(\lambda)$. The eight possible assignments of values to these six quantities leave the set of functions (50) with the values $\{0,1\}$, $\{1,0\}$, $\{\frac{2}{3}, \frac{1}{3}\}$, or $\{\frac{1}{3}, \frac{2}{3}\}$ but never $\{\frac{1}{2}, \frac{1}{2}\}$. This concludes the proof.

B. A proof based on the 2D version of Gleason's theorem

The impossibility of noncontextuality for unsharp measurements and outcome determinism for sharp measurements can also be established in a 2D Hilbert space by making appeal to a recent Gleason-like derivation of the quantum probability rule by Busch [5] and by Caves *et al.* [6]. This "generalized Gleason's theorem" starts from the assumption that there exists a probability measure that assigns a unique probability $w(E)$ to every positive operator E such that $w(I) = 1$, and whenever a set of positive operators forms a resolution of identity, $\Sigma_k E_k = I$, the associated probabilities sum to 1, $\Sigma_k w(E_k) = 1$. From these assumptions, it is proven that the measure must satisfy $w(E) = \mathrm{Tr}(\rho E)$ for some density operator ρ [5,6].

Recall that the values of a set of indicator functions $\{\xi_k(\lambda)\}$ at a particular value of λ form a probability distribution over k. In a measurement noncontextual theory, every positive operator E is represented by a unique indicator function $\xi_E(\lambda)$, with the identity operator being represented by the unit function. Moreover, whenever a set of positive operators forms a resolution of identity, $\Sigma_k E_k = I$, the associated indicator functions sum to the unit function, $\Sigma_k \xi_{E_k}(\lambda) = 1$. Thus the set of all indicator functions for a given value of λ in an ontological model satisfy the assumptions of the set of probability measures in the generalized Gleason's theorem. It follows therefore that for every value of λ in the ontological model, there is a density operator ρ_λ such that $\xi_E(\lambda) = \mathrm{Tr}(\rho_\lambda E)$.

If in addition to measurement noncontextuality, one assumes outcome determinism for sharp measurements, then every projector is represented by a unique idempotent indicator function $\chi_\Pi(\lambda)$, and by the generalized Gleason's theorem,

$$\chi_\Pi(\lambda) = \mathrm{Tr}(\rho_\lambda \Pi). \tag{53}$$

Suppose that $\Pi = |\psi\rangle\langle\psi|$, and consider a λ such that $\chi_\Pi(\lambda) = 1$. In this case, Eq. (53) implies that $\rho_\lambda = |\psi\rangle\langle\psi|$. But then for some other projector $\Pi' = |\psi'\rangle\langle\psi'|$, where $0 < |\langle\psi|\psi'\rangle|^2 < 1$, we have for this value of λ that $\chi_{\Pi'}(\lambda) = \mathrm{Tr}(\rho_\lambda \Pi') = |\langle\psi|\psi'\rangle|^2$ and consequently $0 < \chi_{\Pi'}(\lambda) < 1$,

---

measurement noncontextuality implies that since M and M' are in the same equivalence class, they must be represented by the same set of indicator functions. Thus we require that $\xi_1(\lambda) = \xi_2(\lambda)$. But since $\xi_1(\lambda) + \xi_2(\lambda) = 1$, it follows that $\xi_1(\lambda) = \xi_2(\lambda) = 1/2$ for all λ.

which implies that $\chi_{\Pi'}(\lambda)$ is not idempotent. Thus the assumption of noncontextuality for unsharp measurements and outcome determinism for sharp measurements yields a contradiction in a 2D Hilbert space.

This no-go theorem is related to the no-go theorem of the previous section in the same way that the no-go theorem [1] that is obtained from the standard Gleason's theorem [19] is related to the original Kochen-Specker theorem [2]. The former derive a contradiction using the full set of measurements, while the latter only make use of a finite set.

VI. PROOF OF TRANSFORMATION CONTEXTUALITY IN TWO DIMENSIONS

Consider a set of six transformation procedures, denoted $T_0, T_{\pi/3}, T_{2\pi/3}, T_\pi, T_{4\pi/3}, T_{5\pi/3}$, where the procedure T_θ corresponds to the CP map

$$\mathcal{T}_\theta(\rho) = U_{y,\theta} \rho U_{y,\theta}^\dagger, \tag{54}$$

and where

$$U_{y,\theta} = \begin{pmatrix} \cos\dfrac{\theta}{2} & -\sin\dfrac{\theta}{2} \\ \sin\dfrac{\theta}{2} & \cos\dfrac{\theta}{2} \end{pmatrix} \tag{55}$$

is the unitary operator describing a rotation by θ about the y axis in the Bloch sphere. Consider also the CP map $\mathcal{T}$ that takes all points in the Bloch sphere and projects them onto the $\hat{y}$ axis. There are many ways of implementing $\mathcal{T}$ as a convex sum of transformations, specifically,

$$\mathcal{T} = \frac{1}{2}\mathcal{T}_0 + \frac{1}{2}\mathcal{T}_\pi \tag{56}$$

$$= \frac{1}{2}\mathcal{T}_{\pi/3} + \frac{1}{2}\mathcal{T}_{4\pi/3} \tag{57}$$

$$= \frac{1}{2}\mathcal{T}_{2\pi/3} + \frac{1}{2}\mathcal{T}_{5\pi/3} \tag{58}$$

$$= \frac{1}{3}\mathcal{T}_0 + \frac{1}{3}\mathcal{T}_{2\pi/3} + \frac{1}{3}\mathcal{T}_{4\pi/3} \tag{59}$$

$$= \frac{1}{3}\mathcal{T}_{\pi/3} + \frac{1}{3}\mathcal{T}_\pi + \frac{1}{3}\mathcal{T}_{5\pi/3}. \tag{60}$$

These identities can be explained as follows. The map $\mathcal{T}$ can be achieved by performing with probability $1/2$ a rotation in the Bloch sphere about the y axis by θ and with probability $1/2$ a rotation by $\theta + \pi$. Taking $\theta = 0, \pi/3$, and $2\pi/3$ yields Eqs. (56)–(58). The map $\mathcal{T}$ can also be achieved by performing a rotation about y by θ, by $\theta + 2\pi/3$ or by $\theta + 4\pi/3$ with equal probabilities. Taking $\theta = 0$ and $\pi/3$ yields Eqs. (59) and (60). A rigorous proof of these statements is provided in the Appendix.

By the assumption of transformation noncontextuality each of the seven CP maps we have considered is associated

with a unique transition matrix on the space of ontic states. Suppose that we denote the transition matrix associated with $\mathcal{T}$ by Γ, and the transition matrix associated with $\mathcal{T}_\theta$ by Γ_θ. Because a convex sum of transformation procedures is represented in an ontological model by a convex sum of the associated transition matrices, Eqs. (56)–(60) imply

$$\Gamma = \frac{1}{2}\Gamma_0 + \frac{1}{2}\Gamma_\pi \tag{61}$$

$$= \frac{1}{2}\Gamma_{\pi/3} + \frac{1}{2}\Gamma_{4\pi/3} \tag{62}$$

$$= \frac{1}{2}\Gamma_{2\pi/3} + \frac{1}{2}\Gamma_{5\pi/3} \tag{63}$$

$$= \frac{1}{3}\Gamma_0 + \frac{1}{3}\Gamma_{2\pi/3} + \frac{1}{3}\Gamma_{4\pi/3} \tag{64}$$

$$= \frac{1}{3}\Gamma_{\pi/3} + \frac{1}{3}\Gamma_\pi + \frac{1}{3}\Gamma_{5\pi/3}. \tag{65}$$

Note that $\mathcal{T}_\theta$ and $\mathcal{T}_{\theta+\pi}$ take any rank-1 density operator lying in the z-x plane of the Bloch sphere to a pair of orthogonal density operators. Since these are distinguishable with certainty, it follows from feature 1 of ontological models (see Sec. IV) that the transition matrices Γ_θ and $\Gamma_{\theta+\pi}$ must take any distribution $\mu_x(\lambda)$ associated with such a density operator to disjoint distributions, that is,

$$\int d\lambda' \, \Gamma_\theta(\lambda,\lambda')\mu_x(\lambda') \int d\lambda' \, \Gamma_{\theta+\pi}(\lambda,\lambda')\mu_x(\lambda') = 0. \tag{66}$$

Now consider how our seven transition matrices affect the distribution $\mu_a(\lambda)$ associated with the density operator σ_a, defined in Eq. (14) (recall that σ_a is represented on the Bloch sphere by the vector pointing along the $\hat{z}$ axis). We obtain seven distinct distributions, which we denote $\mu_\theta(\lambda) \equiv \int d\lambda' \, \Gamma_\theta(\lambda,\lambda')\mu_a(\lambda')$, for $\theta = 0, \pi/3, 2\pi/3, \pi, 4\pi/3, 5\pi/3$, and $\mu(\lambda) \equiv \int d\lambda' \, \Gamma(\lambda,\lambda')\mu_a(\lambda')$. By virtue of Eqs. (61)–(65) and Eq. (66), these seven distributions satisfy Eqs. (32)–(36) and Eqs. (24)–(26) where a, A, b, B, c, C are associated with $\theta = 0, \pi, 2\pi/3, 5\pi/3, 4\pi/3, \pi/3$, respectively. But Eqs. (32)–(36) and Eqs. (24)–(26) cannot be satisfied simultaneously, so we have arrived at a contradiction.[11]

VII. IS THE ASSUMPTION OF NONCONTEXTUALITY NATURAL?

An important question is whether the assumption of noncontextuality for preparations, transformations, and unsharp

---

[11]It should be noted that the above argument is equivalent to a proof of preparation contextuality in four dimensions if one makes use of the Jamiolkowski isomorphism between density operators in a 4D space and CP maps in a 2D space [20].

measurements is *as well motivated* as this same assumption for sharp measurements, to which the notion is usually restricted. To answer this, one must consider the motivation for the latter, which seems to be one of ontological economy: be wary of introducing differences in the ontological explanations of empirical phenomena where there are no differences in the phenomena themselves. Einstein's equivalence principle is an example of a fruitful application of this principle. If this is indeed the motivation, then it clearly also applies to our generalized notions of noncontextuality. Specifically, if one believes that equivalent statistics suggest equivalent ontological representations for sharp measurements, why should one not believe this for preparations, transformations, and unsharp measurements as well? Thus, barring an alternative motivation for the traditional notion of noncontextuality, it seems that an ontological model that respects the statistical equivalence class structure of preparations, transformations, and unsharp measurements is as well (or badly) motivated as an ontological model that respects the statistical equivalence class structure of sharp measurements.

This of course leaves open the question of whether *any* assumption of noncontextuality is natural. The answer seems to depend on one's interpretational bent. John Bell, for instance, thought that contextuality was not at all surprising,[12] whereas David Mermin has characterized it as a mystery in need of explanation.[13]

In order to defend the view that measurement contextuality is indeed mysterious within the framework of an ontological model, we show that the reasons for thinking so are very similar to the reasons for thinking that nonlocality is mysterious. Disregarding classical prejudice, nonlocality is not an unreasonable assumption. However, if the universe is fundamentally nonseparable or is such that causal influences can propagate faster than the speed of light, then why should it also be the case that one cannot use these effects to achieve superluminal signalling? Given the presence of nonlocality at the ontological level, it seems almost conspiratorial that one cannot make use of this nonlocality for signalling. Similarly, it is certainly not unreasonable for the statistics of experimental outcomes for a given ontic state to depend on details of the experimental procedure. But assuming this to be the case, it is very surprising that when one considers any valid probability distribution over the ontic states (that is, any distribution that characterizes what someone who knows only the preparation procedure knows about the ontic state), the weighted average over the statistics of outcomes does *not* depend on the details of the experimental procedure. Again, this seems almost conspiratorial. This analogy suggests that removing the appearance of conspiracy from contextuality may well be on a par with reconciling Bell's theorem and relativity as a guide for progress in the search for a wholly

satisfactory realist interpretation of quantum theory.

It is likely that the notion of preparation noncontextuality will also seem natural to some and unnatural to others. To shed some light on the diversity of reactions, it is useful to distinguish two different types of ontological model of quantum theory. Specifically, we distinguish what we call the *epistemic view* and the *ontic view* of quantum states [22].

The epistemic view of quantum states asserts that a density operator represents nothing more than an agent's knowledge about the ontic state of the system. Specifically, it represents the knowledge of someone who knows only the preparation procedure. In this view, the ontic state of a system does not fix the density operator that is used to describe it. Distinct nonorthogonal density operators (including the pure cases) are represented by overlapping probability distributions within this view and are thus consistent with a single ontic state. By contrast, the ontic view of quantum states asserts that the density operator itself represents an attribute of the system, and consequently that two distinct density operators represent mutually exclusive physical states of affairs and are therefore represented in the ontological model by nonoverlapping (i.e., disjoint) probability distributions.

To be precise, for a set $\mathcal{S}$ of density operators (assumed to contain some nonorthogonal elements), an ontological model adopts an ontic view of $\mathcal{S}$ if all distinct elements of $\mathcal{S}$ are represented by disjoint distributions, that is,

$$\rho \neq \rho' \text{ implies } \mu_\rho(\lambda)\mu_{\rho'}(\lambda) = 0 \text{ for all } \rho, \rho' \in \mathcal{S}, \quad (67)$$

whereas an ontological model adopts an epistemic view of $\mathcal{S}$ if only orthogonal elements of $\mathcal{S}$ are represented by disjoint distributions

$$\mu_\rho(\lambda)\mu_{\rho'}(\lambda) = 0 \text{ only if } \rho\rho' = 0, \text{ for all } \rho, \rho' \in \mathcal{S}. \quad (68)$$

In other words, in an epistemic view of $\mathcal{S}$, being orthogonal is a necessary condition for a pair of quantum states to be represented by disjoint distributions (the argument presented at the beginning of Sec. IV shows that orthogonality is a *sufficient* condition for disjointness, regardless of whether one adopts an ontic or an epistemic view).

We now show that an ontic view of the set of pure quantum states rules out the possibility of preparation noncontextuality *trivially*. Our purpose here is to show that an implicit commitment to such a view can lead to the impression that the assumption of preparation noncontextuality is unnatural.

Consider the four preparation procedures P_a, P_A, P_b, and P_B from Sec. IV, represented in quantum theory by the Hilbert space vectors ψ_a, ψ_A, ψ_b, and ψ_B, respectively. An ontic view of pure quantum states implies that not only are the orthogonal states associated with disjoint distributions,

$$\mu_a(\lambda)\mu_A(\lambda) = 0, \quad (69)$$

$$\mu_b(\lambda)\mu_B(\lambda) = 0, \quad (70)$$

but also *nonorthogonal* states are associated with disjoint distributions,

$$\mu_a(\lambda)\mu_b(\lambda) = 0, \quad (71)$$

$$\mu_A(\lambda)\mu_b(\lambda) = 0, \quad (72)$$

---

[12]Bell states: "The result of an observation may reasonably depend not only on the state of the system (including hidden variables) but also on the complete disposition of the apparatus" [1].

[13]Mermin states: "if one is attempting a hidden variable model at all, it seems not unreasonable to expect the model to provide the obvious explanation for this striking insensitivity of the distribution to changes in the experimental arrangement—namely, that the hidden variables are noncontextual" [21].

$$\mu_a(\lambda)\mu_B(\lambda) = 0, \tag{73}$$

$$\mu_A(\lambda)\mu_B(\lambda) = 0. \tag{74}$$

It is then clear that the preparation procedures P_{aA} and P_{bB}, obtained, respectively, by implementing P_a and P_A with equal probability, or P_b and P_B with equal probability, are represented by distributions μ_{aA} and μ_{bB} [defined in Eqs. (27) and (28)] that are also disjoint,

$$\mu_{aA}(\lambda)\mu_{bB}(\lambda) = 0. \tag{75}$$

However, since these two procedures are represented by the same density operator, namely $I/2$, they must be represented by the same distribution in a preparation noncontextual model. Thus an ontic view of quantum states trivially precludes the possibility of preparation noncontextuality.

Since our manner of speaking about pure quantum states typically favors the ontic view of the latter, it also tends to make the assumption of preparation noncontextuality seem implausible. The very term "quantum state" already predisposes one to thinking of the density operator as representing the physical state of affairs rather than an agent's knowledge. For instance, in the context of photon polarization, the multiplicity of convex decompositions of the completely mixed state is sometimes summarized as follows: "an equal mixture of states of horizontal and vertical polarization is statistically indistinguishable from an equal mixture of states of left and right circular polarization." Implicit in this sort of language is the assumption that the four different states of polarization are *mutually exclusive* states of affairs and are therefore ontic states. Indeed, this way of putting things compels us to question (in vain) whether there is not really some measurement that *could* tell these two cases apart. However, it is wrong to take this as an argument against the "naturalness" of preparation noncontextuality because this impression can be attributed entirely to the language that is used to describe the phenomenon.

If one is to take the epistemic view seriously, as one should in an investigation of the possibility of an ontological model of quantum theory, then this sort of language must be avoided, and the assumption of preparation noncontextuality is *a priori* very plausible. Indeed, in light of the arguments that have recently been made in favor of the epistemic view of quantum states [22–25] and the fact that one can reproduce qualitatively many quantum phenomena in noncontextual theories [22,26–28], the impossibility of a preparation noncontextual ontological model appears all the more shocking to the devoted realist.

VIII. THE ISSUE OF OUTCOME DETERMINISM

In our proof of contextuality for unsharp measurements, we assumed outcome determinism for sharp measurements but we assumed outcome *indeterminism* for unsharp measurements. This amounts to representing all and only those POVMs with idempotent elements by sets of indicator functions that are idempotent. Although this seems like a natural assumption to make, two alternative assumptions might seem *a priori* worth considering: (i) that both sharp and unsharp measurements are outcome-deterministic, or (ii) that both are outcome-indeterministic.

We begin by considering the first alternative, that outcome determinism also holds for *unsharp* measurements. It turns out that this is trivially inconsistent with assuming measurement noncontextuality. Consider a measurement procedure M associated with the POVM $\{I/2, I/2\}$. As argued in Sec. V, the assumption of measurement noncontextuality implies that M must be represented in an ontological model by the set of indicator functions $\{1/2, 1/2\}$ which are *not* idempotent, and thus M cannot be outcome deterministic. Recent results by Cabello [4] and Busch [5] also rule out the possibility of a hidden variable model that is measurement noncontextual and outcome deterministic for unsharp measurements. However, these proofs are unnecessarily complex since a consideration of the POVM $\{I/2, I/2\}$ yields the result immediately.

The second alternative is that both sharp and unsharp measurements are outcome-indeterministic. This is the more significant alternative, because it constitutes the weakest assumption and consequently the most general framework for an ontological model. Indeed, unless the assumption of outcome determinism can itself be justified by the assumption of noncontextuality, it is inappropriate to call any no-go theorem that makes use of this assumption a proof of contextuality, because in the face of a contradiction one can always assume that the faulty assumption was that of outcome determinism rather than that of measurement noncontextuality. Thus neither the proof of Bell [1], nor the proof of Kochen and Specker [3], nor any of the proofs of these types including those presented in Sec. V, serve to rule out the possibility of measurement noncontextuality (in the sense in which we have defined the term). It turns out, however, that the assumption of outcome determinism for sharp measurements can be justified by an assumption of *preparation* noncontextuality, as we shall presently demonstrate. Given this inference, the old proofs are vindicated insofar as they remain proofs of the impossibility of *universal* noncontextuality (noncontextuality for all experimental procedures).

It should be noted that Toner, Bacon, and Ben-Or [18] have considered a third alternative, namely, that outcome determinism holds for just those POVMs with elements that are not repeatable, that is, elements that cannot appear twice in a single POVM, and have obtained a nontrivial no-go theorem. Bacciagaluppi [16] has considered a similar alternative and obtained a similar result. Although this is a much weaker assumption than the first alternative, the resulting theorems are still not proofs of the impossibility of universal noncontextuality, according to our definition, since the assumption of outcome determinism for these special POVMs has not been justified by an assumption of universal noncontextuality. In contrast, outcome determinism for all sharp measurements can be so justified. We turn now to the proof of this statement.

A. Preparation noncontextuality implies outcome determinism for sharp measurements

Consider a rank-1 PVM $\{\Pi_k\}$. Thinking of each of the elements as a rank-1 density operator, $\rho_k = \Pi_k$, we obtain an

orthogonal set of rank-1 density operators $\{\rho_k\}$. We denote the density operators and projectors differently because they are represented differently in the ontological model. The set $\{\rho_k\}$ is represented by a set of probability densities $\{\mu_k(\lambda)\}$, while the PVM $\{\Pi_k\}$ is represented by a set of indicator functions $\{\xi_k(\lambda)\}$. Since the ρ_k are orthogonal, the associated preparations are distinguishable with certainty, and thus by feature 1 of ontological models we must have

$$\mu_k(\lambda)\mu_{k'}(\lambda) = \delta_{k,k'}. \tag{76}$$

The support of $\mu_k(\lambda)$, denoted Ω_k, is the region of the ontic state space assigned nonzero probability by $\mu_k(\lambda)$,

$$\Omega_k = \{\lambda | \mu_k(\lambda) > 0\}. \tag{77}$$

Equation (76) then implies that

$$\Omega_k \cap \Omega_{k'} = \emptyset \text{ if } k \neq k'. \tag{78}$$

Now, by virtue of the fact that

$$\text{Tr}(\rho_k \Pi_{k'}) = \delta_{k,k'}, \tag{79}$$

we infer that

$$\int \xi_k(\lambda)\mu_{k'}(\lambda) = \delta_{k,k'}. \tag{80}$$

But, given Eq. (77), this implies that

$$\xi_k(\lambda) = \begin{cases} 1 \text{ for } \lambda \in \Omega_k \\ 0 \text{ for } \lambda \in \cup_{k' \neq k}\Omega_{k'} \end{cases}, \tag{81}$$

or, equivalently,

$$\xi_k(\lambda)\xi_{k'}(\lambda) = \delta_{k,k'} \text{ for } \lambda \in \cup_j \Omega_j. \tag{82}$$

So, if one can show that the union of the supports of the $\mu_k(\lambda)$ is the entire ontic state space, i.e.,

$$\cup_j \Omega_j = \Omega, \tag{83}$$

then Eq. (82) would imply that $\{\xi_k(\lambda)\}$ is a set of idempotent indicator functions, and consequently would establish that our rank-1 PVM must be outcome-deterministic in the ontological model.

It turns out that Eq. (83) follows from the assumption of preparation noncontextuality. First note that the ontic state space Ω can be defined as the set of λ that are assigned nonzero probability by *some* density operator

$$\Omega = \{\lambda | \mu_\rho(\lambda) > 0 \text{ for some } \rho\}. \tag{84}$$

However, since every density operator ρ appears in some convex decomposition of the completely mixed state I/d (where d is the dimensionality of the Hilbert space), and since preparation noncontextuality implies that there is a unique distribution $\mu_{I/d}(\lambda)$ associated with this state, it follows that Ω is simply the set of λ assigned nonzero probability by the latter, i.e.,

$$\Omega = \{\lambda | \mu_{I/d}(\lambda) > 0\}. \tag{85}$$

But given that the ρ_k form a convex decomposition of I/d,

$$\sum_k \frac{1}{d}\rho_k = \frac{I}{d}, \tag{86}$$

it follows from preparation noncontextuality that

$$\sum_k \frac{1}{d}\mu_k(\lambda) = \mu_{I/d}(\lambda), \tag{87}$$

which implies Eq. (83).

This establishes outcome determinism for PVMs all of whose elements are rank 1. Since an arbitrary PVM can always be obtained by coarse-graining of a rank-1 PVM, and since coarse-graining takes idempotent functions to idempotent functions, *any* PVM is represented by a set of idempotent indicator functions. This establishes that the assumption of outcome determinism for sharp measurements follows from an assumption of preparation noncontextuality.

It is natural to wonder whether outcome determinism for sharp measurements might be justified by an assumption of measurement noncontextuality (rather than an assumption of preparation noncontextuality). If this were possible, then the proofs in Sec. V would derive contradictions from measurement noncontextuality alone. It turns out that this is not possible, because measurement noncontextuality on its own is consistent with quantum theory, as we now show.

B. Achieving measurement noncontextuality by giving up outcome determinism

Consider the following ontological model of quantum theory, which is objectively indeterministic and adopts an ontic view of quantum states. The ontic state space Ω is simply taken to be the projective Hilbert space, that is, the set of rays of Hilbert space. Thus for every rank-1 projector $|\psi\rangle\langle\psi|$, we associate a single ontic state, which we denote by ψ. Consequently, there are no hidden variables in this ontological model. A preparation procedure associated with the rank-1 density operator $|\psi'\rangle\langle\psi'|$ is represented by a Dirac-delta distribution

$$\mu_{\psi'}(\psi) = \delta(\psi - \psi'). \tag{88}$$

A preparation procedure involving a convex combination of rank-1 density operators $\{p(\psi'), |\psi'\rangle\langle\psi'|\}$ is represented by the distribution

$$\mu(\psi) = \int d\psi' p(\psi')\delta(\psi - \psi'), \tag{89}$$

where $d\psi$ is the unitarily invariant measure on the projective Hilbert space. A measurement of the POVM $\{Q_k\}$ is associated with a set of indicator functions $\{\xi_{Q_k}(\lambda)\}$ defined by

$$\xi_{Q_k}(\psi) = \text{Tr}(Q_k|\psi\rangle\langle\psi|). \tag{90}$$

These functions are clearly positive by virtue of the positivity of the Q_k, and sum to unity by virtue of the fact that $\Sigma_k Q_k = I$. Note also that they depend only on the POVM that is associated with the measurement and not on how it was implemented. One can see that this model reproduces quantum theory by noting that

$$\int \mu_{\psi'}(\psi)\xi_{Q_k}(\psi)d\psi = \mathrm{Tr}(Q_k|\psi'\rangle\langle\psi'|). \qquad (91)$$

The predictions for mixed preparations are also reproduced.

This model has been discussed at length by Beltrametti and Bugajski [7], and captures to some extent the ontological model that many physicists implicitly adhere to. Note that the model is obviously preparation *contextual* since the distribution that represents a convex combination of preparation procedures, described in Eq. (89), depends on the particular ensemble of pure states, and not just on the density operator associated with the mixture. This fact comes as no surprise since the results of Sec. IV show that *any* ontological model, deterministic or not, must be preparation contextual. More importantly for the purposes of this section, the set of indicator functions associated with any PVM $\{\Pi_k\}$ are not idempotent. This is clear since $\mathrm{Tr}(\Pi_k|\psi\rangle\langle\psi|)$ is only 0 or 1 if $|\psi\rangle$ lies in an eigenspace of Π_k. It follows that the assumption of outcome determinism for sharp measurements is explicitly violated. However, because the set of indicator functions depends only on the POVM, and not on its context, the assumption of measurement noncontextuality is upheld.

IX. CONCLUSIONS

Because the traditional notion of noncontextuality only allowed for a no-go theorem in Hilbert spaces of dimensionality greater than two, there have been many proposed hidden variable models for 2D Hilbert spaces that are purported to be noncontextual [1,3]. These have been presented primarily as pedagogical examples of what sort of model is excluded for larger-dimensional Hilbert spaces. However, by our generalized definition of noncontextuality, all of these models are deemed *contextual* by virtue of being contextual for preparations, transformations, and unsharp measurements. This overturns the notion, suggested by the restriction of old Bell-Kochen-Specker theorems to Hilbert spaces of dimensionality greater than 2, that there is nothing inherently nonclassical about a 2D Hilbert space [29].

In the face of this claim, a skeptic might argue that the proofs presented here have made use of mixed preparations, unsharp measurements, and irreversible transformations (associated, respectively, with non-rank-1 density operators, nonprojective POVMs, and nonunitary CP maps), and that these are necessarily implemented in practice through pure preparations, sharp measurements, and reversible transformations (associated, respectively, with rank-1 density operators, PVMs, and unitary maps) on a larger system and therefore implicitly make use of a Hilbert space of dimension greater than 2. However, this is incorrect. If one examines carefully the proofs presented in this paper, one finds that wherever non-rank-1 density operators, nonprojective POVMs, or nonunitary CP maps arise, they are due to ignorance of which of several rank-1 density operators, PVMs, or unitary maps in the 2D Hilbert space is appropriate, rather than being due to the neglect of a subspace or subsystem of a larger dimensional Hilbert space. In other words, any "ancillary" systems used to implement such procedures can be treated classically, and thus do not require one to posit a larger Hilbert space.

Our operational definition of noncontextuality has allowed us to distinguish the notions of preparation, transformation, and measurement noncontextuality. Our proof of preparation contextuality is particularly interesting as a no-go theorem insofar as it focusses on the impossibility of reproducing, within a particular kind of ontological model, the *convex structure of the set of quantum states* rather than the algebraic structure of the set of quantum measurements. It is interesting to note that wherever one finds a freedom of decomposition in the formalism of operational quantum theory, such as the multiplicity of convex decompositions of a mixed quantum state or of a POVM element, the multiplicity of fine-grainings of a non-rank-1 POVM, or the unitary freedom in the operator-sum representation of a nonunitary CP map, one can develop a proof of contextuality that is based on this freedom.

We have shown that one can confine all the contextuality into the preparations and transformations if one likes, because there exist outcome-indeterministic ontological models of quantum theory, such as the Beltrametti-Bugajski model, that are measurement noncontextual. On the other hand, one cannot confine all the contextuality into the measurements, because the assumption of preparation noncontextuality yields a contradiction on its own. In this sense, preparation contextuality is more fundamental to quantum theory than measurement contextuality.

The issue of noncontextuality is closely linked with the issue of locality. Indeed, it is sometimes claimed that nonlocality is an instance of measurement contextuality. If this were the case, then proofs of nonlocality would also constitute proofs of measurement contextuality, and since there exist proofs of nonlocality that do not assume outcome determinism for sharp measurements, it would appear that there should exist proofs of measurement contextuality that do not make this assumption either. But this would be in contradiction with the claims of the previous section.

The resolution of this puzzle is that one can distinguish two sorts of locality [30], and it is only the failure of one of these that implies measurement contextuality. The first notion of locality, which we call *separability*, is the assumption that the ontic state of the universe is defined in terms of the ontic states at each point of space-time. The other sort of locality assumption, which presumes separability, we call *local causality*. It is the assumption that the probability distribution over values for a variable in a space-time region are determined by the values of all the variables in the backward light cone of this region (see footnote in Sec. III). A failure of local causality within the framework of a separable model *does* indeed imply measurement contextuality. However, a model can be nonlocal by virtue of failing to be separable, and in this case it does not follow that the model is measurement contextual. This is precisely what occurs in the Beltrametti-Bugajski model. The variables for a composite system are not simply the Cartesian product of the variables of the components, since the Cartesian product of two projective Hilbert spaces is not the projective Hilbert space of the tensor product (it fails to include the entangled states). In particular, spatially separated systems are not associated with distinct variables. Thus the Beltrametti-Bugajski model is not separable. It is only within the context of a separable theory

that Bell's theorem implies measurement contextuality.

The opposite inference has also received a great deal of attention: whether a proof of measurement contextuality can be turned into a proof of nonlocality (note that the question is only interesting if one presumes separability since otherwise one is already acknowledging a failure of some sort of locality). The motivation for this investigation is clear since, as Bell famously emphasized, an assumption of measurement noncontextuality is most compelling if it can be justified by an assumption of locality [31]. Many authors have shown how certain no-go theorems for measurement noncontextuality can be turned into no-go theorems for locality by virtue of the fact that sometimes every assumption of measurement noncontextuality in a Bell-Kochen-Specker theorem can be justified by an assumption of locality [21,32]. It turns out that the same trick can be achieved in no-go theorems for preparation noncontextuality. Although the particular proof of the no-go theorem presented in Sec. IV does not admit such a justification, a proof can be found which does. This will be presented in a separate paper [33]. The version of Bell's theorem that results is particularly enlightening, as it constitutes a more direct response to the Einstein, Podolsky, and Rosen argument [34] compared to standard versions of the theorem.

It should be noted that there are contexts that do not have any representation in the formalism of operational quantum theory. Whether one uses a piece of polaroid or a birefringent crystal in a measurement of photon polarization is an example of such a context. No dependence on *this sort of context* is implied by any of the no-go theorems we have presented.[14] Nonetheless, some hidden variable theories still exhibit such dependence. For instance, it has been shown that the deBroglie-Bohm interpretation has this sort of context dependence for certain position measurements [35,36] and for certain spin measurements [37]. Thus the deBroglie-Bohm interpretation involves more contextuality than has been shown to be required of an ontological model. Note that Refs. [35–37] explicitly identify this feature of the deBroglie-Bohm interpretation as a kind of contextuality despite the fact that it does not fit into the standard definition of contextuality presented in the Introduction. The possibility of this type of phenomenon was in fact considered in the framework of a general hidden variable theory much earlier by Shimony [38], who also described it as a kind of contextuality. This highlights another virtue of our generalized definition of contextuality: it accords with the intuition that a measurement context is *any* feature of the measurement that is not specified by specifying its equivalence class.

An operational definition of noncontextuality is also likely to be useful because it allows one to investigate the possibility of finding ways of experimentally differentiating the set of noncontextual theories from the set of contextual theories, much as the Bell inequalities differentiate all local

realistic theories from their alternatives. If these investigations are successful, they could shed light on the question of how to perform experimental tests of contextuality, a subject of much recent interest [39–41]. The question of whether an experimental test of contextuality is even *possible* has been the subject of some controversy, due to the finite precision of real experimental procedures [42–45]. The problem, from the perspective of this paper, is that finite precision might imply that in practice no two experimental procedures are found to be operationally equivalent, in which case the assumption of noncontextuality is never applicable. A possible resolution of this finite precision loophole is to further generalize the definition of noncontextuality proposed in the introduction as follows:

A noncontextual ontological model of an operational theory is one wherein if two experimental procedures are operationally similar, then they have similar representations in the ontological model.

To be substantive, this proposal must be supplemented by a quantitative measure of similarity in the space of operational procedures, and a corresponding measure in the space of ontological representations of these procedures. Whether this strategy can lead to an experimentally robust notion of contextuality is a subject for future research.

Finally, given the fact that some quantum information processing protocols, namely, protocols for communication complexity problems [46], have been proven to require violations of Bell inequalities in order to outperform their classical counterparts, it is interesting to investigate whether the power of any quantum information processing protocols might be attributed to the contextuality of quantum theory. There is already some evidence to this effect in the case of random access codes [47]. We speculate that this might also be the case for the exponential speedup of a quantum computer relative to a classical computer, if such a speedup exists.

---

[14]If, however, one decides to treat part of the experimental apparatus as a quantum system, then this sort of distinction *could* be represented within the quantum formalism and it is then conceivable that one could prove that context dependence of this type is sometimes required.

APPENDIX: PROOF OF EQS. (56)–(60)

To demonstrate Eqs. (56)–(60), we make use of the fact that there is unitary freedom in the operator-sum representation of a CP map [48]. Suppose that $\{W_\mu\}$ are a set of opera-

tors (called *Kraus operators*) appearing in an operator-sum representation of $\mathcal{T}$, that is,

$$\mathcal{T}(\rho) = \sum_\mu W_\mu \rho W_\mu^\dagger. \tag{A1}$$

Then, for any unitary matrix $u_{\nu\mu}$, the set of operators $\{X_\nu\}$ defined by

$$X_\nu = \sum_\mu u_{\nu\mu} W_\mu \tag{A2}$$

also forms an operator-sum representation of $\mathcal{T}$. Note that we allow Kraus operators to be zero, so that different operator-sum representations may have different cardinality.

Equation (56) implies that $\mathcal{T}$ has an operator-sum representation in terms of the set of Kraus operators $\{W_1, W_2\} = \{(1/\sqrt{2})U_0, (1/\sqrt{2})U_\pi\}$, since

$$\mathcal{T}(\rho) = \frac{1}{2} U_0 \rho U_0^\dagger + \frac{1}{2} U_\pi \rho U_\pi^\dagger. \tag{A3}$$

The set of operators $\{X_1, X_2\} = \{(1/\sqrt{2})U_\theta, (1/\sqrt{2})U_{\theta+\pi}\}$ also yield an operator-sum representation of $\mathcal{T}$ since they can be obtained by a unitary remixing of $\{W_1, W_2\}$, via Eq. (A2), using the 2×2 unitary matrix

$$u = \begin{pmatrix} \cos\dfrac{\theta}{2} & \sin\dfrac{\theta}{2} \\ -\sin\dfrac{\theta}{2} & \cos\dfrac{\theta}{2} \end{pmatrix}. \tag{A4}$$

It follows, in particular, that the sets of Kraus operators $\{(1/\sqrt{2})U_{\pi/3}, (1/\sqrt{2})U_{4\pi/3}\}$ and $\{(1/\sqrt{2})U_{2\pi/3}, (1/\sqrt{2})U_{5\pi/3}\}$ form operator-sum representations of $\mathcal{T}$, and consequently that Eqs. (57) and (58) hold.

Next, we show that the set of operators $\{X_1, X_2, X_3\} = \{(1/\sqrt{3})U_\theta, (1/\sqrt{3})U_{\theta+2\pi/3}, (1/\sqrt{3})U_{\theta+4\pi/3}\}$ also yield an operator-sum representation of $\mathcal{T}$. First note that the set $\{W_1, W_2, W_3\} = \{(1/\sqrt{2})U_0, (1/\sqrt{2})U_\pi, 0\}$ yields the operator-sum representation of $\mathcal{T}$ associated with Eq. (56). The operators $\{X_1, X_2, X_3\}$ can be obtained by a unitary remixing of $\{W_1, W_2, W_3\}$ using the 3×3 unitary matrix

$$u = \begin{pmatrix} \sqrt{\dfrac{2}{3}}\cos\dfrac{\theta}{2} & \sqrt{\dfrac{2}{3}}\sin\dfrac{\theta}{2} & \sqrt{\dfrac{1}{3}} \\ \sqrt{\dfrac{2}{3}}\cos\left(\dfrac{\theta}{2}+\dfrac{2\pi}{3}\right) & \sqrt{\dfrac{2}{3}}\sin\left(\dfrac{\theta}{2}+\dfrac{2\pi}{3}\right) & \sqrt{\dfrac{1}{3}} \\ \sqrt{\dfrac{2}{3}}\cos\left(\dfrac{\theta}{2}+\dfrac{4\pi}{3}\right) & \sqrt{\dfrac{2}{3}}\sin\left(\dfrac{\theta}{2}+\dfrac{4\pi}{3}\right) & \sqrt{\dfrac{1}{3}} \end{pmatrix}. \tag{A5}$$

It follows, in particular, that $\{(1/\sqrt{3})U_0, (1/\sqrt{3})U_{2\pi/3}, (1/\sqrt{3})U_{4\pi/3}\}$ and $\{(1/\sqrt{3})U_{\pi/3}, (1/\sqrt{3})U_\pi, (1/\sqrt{3})U_{5\pi/3}\}$ form operator-sum representations of $\mathcal{T}$ and consequently that Eqs. (59) and (60) hold.

---

[1] J. S. Bell, Rev. Mod. Phys. **38**, 447 (1966), reprinted in Ref. [2], Chap. 1.

[2] J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics* (Cambridge University Press, New York, 1987).

[3] S. Kochen and E. P. Specker, J. Math. Mech. **17**, 59 (1967).

[4] A. Cabello, Phys. Rev. Lett. **90**, 190401 (2003).

[5] P. Busch, Phys. Rev. Lett. **91**, 120403 (2003).

[6] C. M. Caves, C. A. Fuchs, K. Manne, and J. M. Renes, Found. Phys. **34**, 193 (2004).

[7] E. G. Beltrametti and S. Bugajski, J. Phys. A **28**, 3329 (1995).

[8] D. Bohm, Phys. Rev. **85**, 166 (1952).

[9] A. Peres, *Quantum Theory: Concepts and Methods* (Kluwer Academic, Boston, 1995).

[10] L. Hardy, e-print quant-ph/0101012.

[11] K. Kraus, *States, Effects and Operations: Fundamental Notions of Quantum Theory*, Lecture Notes in Physics No. 190 (Springer-Verlag, Berlin, 1983).

[12] P. Busch, M. Grabowski, and P. Lahti, *Operational Quantum Physics* (Springer, Berlin, 1995).

[13] B. Schumacher, Phys. Rev. A **54**, 2614 (1996).

[14] A. Peres, Found. Phys. **20**, 1441 (1990).

[15] J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics* (Ref. [2]), Chap. 7.

[16] G. Bacciagaluppi (unpublished).

[17] R. W. Spekkens and T. Rudolph, Quantum Inf. Comput. **2**, 66 (2002).

[18] B. F. Toner, D. Bacon, and M. Ben-Or (unpublished).

[19] A. M. Gleason, J. Math. Mech. **6**, 885 (1957).

[20] A. Jamiolkowski, Rep. Math. Phys. **3**, 275 (1972).

[21] N. D. Mermin, Rev. Mod. Phys. **65**, 803 (1993).

[22] R. W. Spekkens, e-print quant-ph/0401052.

[23] L. E. Ballentine, Rev. Mod. Phys. **42**, 358 (1970); L. E. Ballentine, Y. Yang, and J. P. Zibin, Phys. Rev. A **50**, 2854 (1994).

[24] C. A. Fuchs, e-print quant-ph/0205039; J. Mod. Opt. **50**, 987 (2003).

[25] J. V. Emerson, Ph.D. thesis, Simon Fraser University, Vancouver, Canada, 2001; e-print quant-ph/0211035.

[26] L. Hardy, e-print quant-ph/9906123.

[27] K. A. Kirkpatrick, Found. Phys. Lett. **16**, 199 (2003).

[28] T. Rudolph (unpublished).

[29] S. J. van Enk, Phys. Rev. Lett. **84**, 789 (2000).

[30] D. Howard, Stud. Hist. Philos. Sci **16**, 171 (1985).

[31] J. S. Bell, Physics (Long Island City, N.Y.) **1**, 195 (1964), reprinted in Ref. [2], Chap. 2.

[32] J. Zimba and R. Penrose, Stud. Hist. Philos. Sci **24**, 697 (1993).

[33] T. Rudolph and R. W. Spekkens (unpublished).

[34] A. Einstein, B. Podolsky, and N. Rosen, Phys. Rev. **47**, 777

(1935).

[35] A. Valentini, Ph.D. thesis, International School for Advanced Studies, Trieste, Italy, 1991.

[36] L. Hardy, in *Bohmian Mechanics and Quantum Theory: An Appraisal*, edited by D. J. T. Cushing, A. Fine, and S. Goldstein (Kluwer Academic, New York, 1996), pp. 67–76.

[37] D. Z. Albert, *Quantum Mechanics and Experience* (Harvard University Press, Cambridge, England, 1992), p. 153.

[38] A. Shimony, Br. J. Philos. Sci. **35**, 25 (1984).

[39] C. Simon, C. Brukner, and A. Zeilinger, Phys. Rev. Lett. **86**, 4427 (2001).

[40] J.-A. Larsson, Europhys. Lett. **58**, 799 (2002).

[41] A. Cabello and G. Garcia-Alcaine, Phys. Rev. Lett. **80**, 1797 (1998).

[42] D. A. Meyer, Phys. Rev. Lett. **83**, 3751 (1999).

[43] R. Clifton and A. Kent, Proc. R. Soc. London, Ser. A **456**, 2101 (2000).

[44] D. M. Appleby, Phys. Rev. A **65**, 022105 (2002).

[45] J. Barrett and A. Kent, Stud. Hist. Philos. Mod. Phys. **35**, 151 (2004).

[46] C. Brukner, M. Zukowski, J.-W. Pan, and A. Zeilinger, Phys. Rev. Lett. **92**, 127901 (2004).

[47] E. F. Galvão, Ph.D. thesis, University of Oxford, Oxford, England; e-print quant-ph/0212124.

[48] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, England 2000), p. 372.