Information processing in generalized probabilistic theories

Jonathan Barrett[*]
Perimeter Institute for Theoretical Physics, 31 Caroline Street N, Waterloo, Ontario N2L 2Y5, Canada

(Received 25 May 2006; published 5 March 2007)

I introduce a framework in which a variety of probabilistic theories can be defined, including classical and quantum theories, and many others. From two simple assumptions, a tensor product rule for combining separate systems can be derived. Certain features, usually thought of as specifically quantum, turn out to be generic in this framework, meaning that they are present in all except classical theories. These include the nonunique decomposition of a mixed state into pure states, a theorem involving disturbance of a system on measurement (suggesting that the possibility of secure key distribution is generic), and a no-cloning theorem. Two particular theories are then investigated in detail, for the sake of comparison with the classical and quantum cases. One of these includes states that can give rise to arbitrary nonsignaling correlations, including the superquantum correlations that have become known in the literature as nonlocal machines or Popescu-Rohrlich boxes. By investigating these correlations in the context of a theory with well-defined dynamics, I hope to make further progress with a question raised by Popescu and Rohrlich, which is why does quantum theory not allow these strongly nonlocal correlations? The existence of such correlations forces much of the dynamics in this theory to be, in a certain sense, classical, with consequences for teleportation, cryptography, and computation. I also investigate another theory in which all states are local. Finally, I raise the question of what further axiom(s) could be added to the framework in order to identify quantum theory uniquely, and hypothesize that quantum theory is optimal for computation.

DOI: 10.1103/PhysRevA.75.032304 PACS number(s): 03.67.−a, 03.65.Ta

I. INTRODUCTION

A question periodically raised is what is responsible for the power of quantum computation (or cryptography, or information processing in general)? At a recent meeting in Konstanz [1], speakers referred to quantum entanglement; the superposition principle; the exponentially growing size of Hilbert space with the number of qubits; nonlocality and contextuality; the possibility of continuous reversible transformations between pure states; and what is known as the sign problem in Monte Carlo simulations of certain types of quantum systems [2]. It is perhaps unsurprising that there are so many different answers. The problem is that the results of quantum information theory are already well understood as consequences of the quantum formalism, and it is not clear that simply pointing to aspects of that formalism tells us anything new. What we are really looking for is a better understanding of the connections between information processing and physical principles in general.

Such an understanding could be gained by studying information processing in a broader range of theories than classical and quantum, where different physical principles may hold. For any theory, whether it applies to Nature or not, one can consider the information processing possibilities of this theory, the differences from those of classical or quantum theory, and attempt to trace these possibilities back to the fundamental features of the theory. Some authors have indeed investigated unrealistic theories, with a view to understanding the relevant features [3–13].

To make further progress along these lines, I introduce an operational framework for probabilistic theories in which a broad range of different theories can be defined. The framework, described in Secs. II–IV, is based on that used by Hardy in his derivation of quantum theory from simple axioms [14]. The basic idea is that a state is represented as a vector of probabilities of measurement outcomes. Transformations of a system must correspond to linear transformations of this vector. By including probabilistic, that is normalization-decreasing, transformations, a unified account of transformations and measurements can be given. Rather than employ any of Hardy's axioms, I introduce two assumptions that concern how separate systems combine to form a joint system. The first is that operations on the separate systems commute (this implies a no-signaling principle), and the second is that the state of the joint system can be completely specified by joint probabilities for local measurements. From these assumptions a tensor product rule can be derived. This removes at least some of the mystery from the quantum tensor product rule and generalizes a derivation by Fuchs [15].

The resulting framework includes classical probabilistic theories, quantum theory, and many other theories besides. The first thing one notices is that certain phenomena, usually thought of as specifically quantum, are in fact generic. This means that they either appear in all theories, or they appear in all theories except classical theories, which emerge as a very special case. As shown in Sec. V, these phenomena include the nonunique decomposition of a mixed state into pure states, a theorem concerning the disturbance of a system on measurement, and the no-cloning theorem. (These observations are complementary to those of Ref. [13], where it is noted that similar properties hold in nonlocal but nonsignaling theories.)

In addition to looking at generic properties of theories, it is useful to analyze at least one or two interesting theories in detail. These then provide well-understood examples that can be contrasted with the classical and quantum cases. Thus the

[*] Electronic address: jbarrett@perimeterinstitute.ca

rest of this work is devoted to an analysis of two theories that admit a particularly natural definition. The first of these allows arbitrary correlations between measurements on separated systems, as long as they are nonsignaling. I call it generalized nonsignaling theory (GNST). The correlations allowed by this theory can be more nonlocal than quantum theory allows, and include the superquantum correlations that have come to be known variously in the literature as Popescu-Rohrlich (PR) boxes, or nonlocal machines [16–26]. Popescu and Rohrlich raised the question of why quantum theory does not allow these correlations. An investigation of a complete theory, with dynamics, that does include the correlations may help to answer this question. The second theory allows the same states of single systems as GNST, but does not allow any violation of Bell inequalities. For this reason it is called generalized local theory (GLT).

One of the interesting things about GNST is that there are many direct analogs of quantum phenomena (in addition to the generic phenomena mentioned above). These include entanglement, nonlocality, a form of contextuality, and the Einstein-Podolski-Rosen (EPR) paradox. (Interestingly, a quite different toy theory introduced by Spekkens displays many of these phenomena too [11].) However, there are also differences with quantum theory. A central insight of this work is that there is a tradeoff between the allowed states of a theory and the allowed dynamics. This follows from the simple fact that dynamics has to act in such a way that allowed states are taken to allowed states. In the case of GNST, the fact that all nonsignaling correlations are possible means that the dynamics is highly restricted. In fact, I show in Sec. VI that the dynamics of single systems in GNST is essentially classical, corresponding to no more than relabelings of measurements and outcomes. This result is extended to transformations and measurements on simple kinds of bipartite systems (more complicated cases are still open). GLT is in some sense intermediate, with transformations on single systems similarly simple, but with transformations on bipartite systems including other possibilities.

These conclusions about dynamics have consequences for information processing, discussed in Sec. VII. For example, there is no teleportation in GNST, despite the existence of highly nonlocal states that might have been thought to facilitate a task like teleportation. Key distribution is possible in GNST and 1-2 oblivious transfer in both GNST and GLT. Other cryptographic possibilities, such as key distribution in GLT, or bit commitment in either theory, are open questions. A natural circuit-type model of computation can be defined for any theory in the framework. The states and dynamics together in GLT are sufficiently restricted that computation can be simulated efficiently by a classical computer. The theorems concerning dynamics in GNST give evidence that computation in this theory can also be simulated efficiently by a classical computer (*despite* the existence of superentangled states). The fact that quantum theory, unlike GNST and GLT, achieves such a harmonious balance of states and dynamics leads to the following hypothesis that I leave open: *a quantum computer can simulate computation in any theory in the framework with at most polynomial overhead*.

Finally, two motivations are not directly connected with information processing. On the face of it, most of the theories that can be written down in the framework described suffer from similar interpretational problems as quantum theory. For example, are pure states in one of these theories best regarded as complete descriptions of individual reality, as describing only ensembles, or as descriptions of agents' degrees of belief? Although I do not do this in this paper, consideration of these questions in a broader framework may shed new light on the quantum theoretical problems. The other motivation is to stimulate research into finding ways of deriving quantum theory from physical principles (instead of laying down a list of mathematical axioms, as per the standard textbook approach). What principles could be used to rule out the other theories described and leave only quantum theory? One reason for deriving quantum theory from physical principles is that by modifying one or another of the principles, we may discover new ways of going beyond quantum theory.

II. A FRAMEWORK FOR PROBABILISTIC THEORIES

This section describes in some detail a general operational framework in which probabilistic theories can be written down. All theories in this framework share the following features with classical and quantum theory.

(i) Local operations on distinct subsystems commute. In the case of a bipartite system AB, for example, this means that if an operation is performed on system A alone, and an operation on system B alone, it does not matter what order the operations were performed in.

(ii) The global state of a composite system is determined by correlations between local measurements.

A. States and operations

Consider a laboratory containing *preparation devices* and *operation devices*. Preparation devices prepare a system in a given state and operation devices act on a system, in general changing its state. When an operation device is used, there may be several different outcomes, each occurring with some probability. Each outcome is identified by a different macroscopic event (for example, a different light being illuminated on the device, or a different position of a pointer). Thus operation devices serve to perform both transformations and measurements. Given the state of a system, it should be possible to calculate the probabilities of measurement outcomes for any measurement. Conversely, if the probabilities of measurement outcomes for any measurement are known, then the state is known.

Suppose that systems come in different *types*, where in quantum theory, for example, the type of system corresponds to the dimension of its Hilbert space. For each type of system, there is some finite set $\mathcal{F}$ of measurements, each with a finite number of outcomes, such that the state of the system can be completely specified by listing the probabilities for these outcomes. For example, in quantum theory, the state of a spin-$1/2$ particle can be specified by giving the probabilities of obtaining spin-up on measuring in the x, y, and z directions. Call the measurements in $\mathcal{F}$ *fiducial* measurements and $\mathcal{F}$ the *fiducial set*. In general, there will be other measurements that can be performed on a system that are not contained in the fiducial set (a measurement of spin in some direction at 45° to the z axis, say). The probabilities of outcomes of these measurements can nevertheless be determined from the state. We ignore the possibility of states requiring an infinite number of probabilities to be specified (despite the fact that quantum theory includes infinite dimensional systems and classical probability theory infinite sample sets). The set of fiducial measurements need not be unique. In general it will be possible to find a different set (perhaps involving a different number of measurements with different numbers of outcomes) that also suffices to specify the state.

This is essentially the framework described by Hardy [14], who introduced the term *fiducial* for the state-defining measurements. (See also [15,27–29], where the idea of representing a state via probabilities for measurement outcomes is also explored.) Unlike Hardy, we shall assume for convenience that the degrees of freedom expressed in the state are internal degrees of freedom, and that all measurements are measurements of internal degrees of freedom. With respect to spacetime degrees of freedom, systems behave classically, having a definite position and velocity at all times. This seems the most natural position to take given that we are most interested in the information processing properties of the different theories considered. However, it would be interesting to extend this work, and to consider what Nature would be like if all degrees of freedom, including those of spacetime, were described by a theory like one of the ones presented here (but extended to allow for infinite-outcome measurements).

The above is summarized by the following.

Assumption 1. The state of a single system can be completely specified by listing the probabilities for the outcomes of some subset $\mathcal{F}$ of all possible measurements. These are the fiducial measurements. These probabilities can be written arranged in a vector,

$$\vec{P} \equiv \left(\begin{array}{c} P(a=1|X=1) \\ P(a=2|X=1) \\ \vdots \\ \hline P(a=1|X=2) \\ P(a=2|X=2) \\ \vdots \\ \hline \vdots \end{array} \right). \qquad (1)$$

$P(a{=}i|X{=}j)$ is the probability of getting outcome i when fiducial measurement $j \in \mathcal{F}$ is performed on the system.

Normalization of the state would require that

$$\sum_i P(a=i|X=j) = 1 \quad \forall \;\; j, \qquad (2)$$

where the sum ranges over all the values i that the outcome can take for a particular measurement. It is convenient also to give a meaning to unnormalized states (just as in quantum theory it is sometimes convenient to write down unnormalized density matrices). Suppose that a system is prepared in some (normalized) state and an operation performed with an outcome i that is obtained with probability less than 1. There is an unnormalized state associated with i, each entry of which is the joint probability of getting i followed by a particular outcome for a subsequent fiducial measurement. This implies that unnormalized states satisfy

$$\sum_{i'} P(a=i'|X=j) = \sum_{i''} P(a=i''|X=j') = c \quad \forall \;\; j,j' \qquad (3)$$

with $0 \leq c \leq 1$. In the case described, c is the probability of the outcome i. This idea generalizes to chains of operations, thus operations should be defined on unnormalized states as well as on normalized ones. Define

$$|\vec{P}| \equiv \sum_i P(a=i|X=j), \qquad (4)$$

where the right-hand side is independent of the choice of j. The notation $|\vec{P}|$ is used throughout and should not be confused with more usual definitions of the norm of a vector.

Suppose that for each type of single system, the fiducial measurements are fixed. A particular theory will specify, for each type of system, a set of allowed vectors $\vec{P}$. These correspond to physically possible states of a system, i.e., states that can actually be prepared using one of the preparation devices. There is no reason to suppose that all vectors $\vec{P}$ that can be written down can actually be prepared. For example, in quantum theory, one cannot prepare a system that will with certainty return the outcome spin-up for spin measurements in both the z and x directions. Call the set of allowed states $\mathcal{S}$ (where there is a different $\mathcal{S}$ for each type of system but we suppress this dependence).

Assumption 2. For each type of system, the set of allowed normalized states is closed and convex. The complete set of states $\mathcal{S}$ is the convex hull of the set of allowed normalized states and $\vec{0}$.

$\vec{0}$ is the vector with all entries 0. The idea behind this assumption is that it is always possible to toss a biased coin and subsequently to be interested only in the joint probabilities of getting given measurement outcomes along with heads. In this way one can "prepare" unnormalized states. If heads occurs with probability zero, the state $\vec{0}$ is prepared. Convexity of $\mathcal{S}$ corresponds to the assumption that if it is possible to prepare states P_1 and P_2, then it is also possible to prepare any probabilistic mixture of the two states. One may toss a coin, prepare either P_1 or P_2 depending on the outcome, and then forget the outcome.[1] Extreme points of $\mathcal{S}$ apart from $\vec{0}$ are *pure states*. States that are neither pure nor $\vec{0}$ are *mixed*. Mixed states can be written as a convex sum of pure states and $\vec{0}$, but this sum need not be unique.

Notice from Eq. (3) that $\mathcal{S}$ lies in a subspace of the complete vector space. In general, we allow for the possibility that $\vec{P}$ is an overcomplete description of the state of a system. Thus there may be other linear constraints that apply apart

[1] The assumption is also stated in such a manner as to rule out the possibility of an unnormalized state without a corresponding normalized state.

from Eq. (3) implying that $\mathcal{S}$ lies in a smaller subspace still.

When an operation is performed, each outcome is associated with a transformation of the state of the system, i.e., with a map from states to states:

$$\vec{P} \to \vec{P}' = f(\vec{P}). \tag{5}$$

Some operations have only one outcome and the corresponding transformation preserves normalization of the state (in quantum theory, these are the trace-preserving completely positive maps). If an outcome occurs with probability <1, then it is associated with a transformation that decreases the normalization of the state (in quantum theory, these are trace-decreasing completely positive maps). In the most general case, one could consider operations that change the system into a system of a different type (just as in quantum theory one sometimes considers completely positive maps between Hilbert spaces of different dimension). In this work I assume that operations do not change the type of system, although the appropriate generalization is not usually too difficult.

Consider a transformation acting on a system that is in a mixed state, that is a state $\vec{P}$ such that

$$\vec{P} = \sum_i q_i \vec{P}_i, \tag{6}$$

where the $\vec{P}_i$ are allowed states and where $0 \leqslant q_i \leqslant 1$ and $\Sigma_i q_i = 1$. One way of preparing a system in such a state would be to prepare a system in the state $\vec{P}_i$ with probability q_i and then to forget the value of i. In this case the transformed $\vec{P}$ must be the same convex combination of the transformed $\vec{P}_i$, that is

$$f(\vec{P}) = f\left(\sum_i q_i \vec{P}_i\right) = \sum_i q_i f(\vec{P}_i) \quad \forall \;\; P_i \in \mathcal{S}. \tag{7}$$

It follows from this that the action of f on the set of allowed states $\vec{P}$ can be represented as

$$\vec{P} \to M \cdot \vec{P}, \tag{8}$$

where M is a matrix, i.e., f is a linear map. This is not completely obvious from Eq. (7), since the equation involves only convex combinations, and furthermore only applies for those $\vec{P}_i \in \mathcal{S}$. A rigorous proof is given in Appendix A.

An operation corresponds to a set of matrices $\{M_i\}$.[2] The unnormalized state associated with the ith outcome is $M_i \cdot \vec{P}$, and the unnormalized probability of the ith outcome is

$$|M_i \cdot \vec{P}|. \tag{9}$$

For each type of system, a particular theory will specify a set of allowed operations. Denote this set $\mathcal{O}$. An element of $\mathcal{O}$ is a set of transformations $\{M_i\}$, and must be such that the following holds.

Constraint 1.

$$0 \leqslant \frac{|M_i \cdot \vec{P}|}{|\vec{P}|} \leqslant 1 \quad \forall \;\; i, \vec{P} \in \mathcal{S}, \tag{10}$$

$$\sum_i \frac{|M_i \cdot \vec{P}|}{|\vec{P}|} = 1 \quad \forall \;\; \vec{P} \in \mathcal{S}, \tag{11}$$

$$M_i \cdot \vec{P} \in \mathcal{S} \quad \forall \;\; i, \vec{P} \in \mathcal{S}. \tag{12}$$

A further constraint is that each transformation M_i must result only in allowed states when it acts on a system that is part of a larger multipartite system (see the next section). The following assumption results in some loss of generality but also makes things simpler.

Assumption 3. For each type of system, there is a set $\mathcal{T}$ of allowed transformations. A set of transformations $\{M_i\}$ is an element of $\mathcal{O}$ if and only if $M_i \in \mathcal{T} \forall \;\; i$, and Eq. (11) is satisfied. The set $\mathcal{T}$ includes the transformation that maps all $\vec{P}$ to $\vec{0}$ and is convex.

With this assumption, once $\mathcal{T}$ is given, a separate specification of $\mathcal{O}$ is not needed. The reasons for convexity are similar to those given for Assumption 2.

As mentioned above, the formalism of operations already includes measurements. Sometimes one is not interested in the state after measurement but only in the probabilities of the different outcomes. In this case it is convenient to associate with an operation $\{M_i\}$ a set of vectors $\{R_i\}$ such that

$$\vec{R}_i \cdot \vec{P} = |M_i \cdot \vec{P}| \quad \forall \;\; \vec{P} \in \mathcal{S}. \tag{13}$$

Such a set can always be found. For a normalized $\vec{P}$, the probability of the ith outcome is then given by $\vec{R}_i \cdot \vec{P}$. It does not matter if the vector $\vec{R}_i$ is not unique—this simply means that different vectors can represent the same measurement outcome. Denote by $\mathcal{M}$ the set of all sets $\{R_i\}$ such that Eq. (13) holds for some $\{M_i\} \in \mathcal{O}$. $\mathcal{M}$ is the set of allowed measurements. Denote by $\mathcal{R}$ the set of allowed measurement vectors, that is, the set of vectors $\vec{R}$ such that $\vec{R} \cdot \vec{P} = |M \cdot \vec{P}| \forall \;\; \vec{P} \in \mathcal{S}$, for some $M \in \mathcal{T}$.[3] (Notation: $\mathcal{R}$ should not be confused with $\mathbb{R}$, the set of real numbers.)

[2] A note on terminology. I shall continue to use the term *operation* to refer to the experiment with a number of different outcomes corresponding to the set $\{M_i\}$, and the term *transformation* to refer to a single, in general normalization-decreasing, M_i.

[3] Recall that in quantum theory, an effect E is a positive operator such that $0 \leqslant E \leqslant 1$. $\vec{R}$ vectors are essentially a generalization of the effects to our framework. In the usual quantum formalism, an effect can represent an outcome of a measurement on a quantum state ρ, with the probability of the outcome given by $\mathrm{Tr}(E\rho)$. A set of effects E_i such that $\Sigma_i E_i = I$, where I is the identity, is a positive operator-valued (POV) decomposition of the identity, and corresponds to a POV measurement.

B. Multipartite systems

So far, the framework described is similar to that used by Hardy as a starting point for his derivation of quantum theory (although I have been more explicit about treating transformations and measurements in a unified manner). Hardy narrows things down with various axioms. Rather than adopt any of Hardy's axioms, however, I introduce a small number of nontrivial assumptions that concern how systems combine to make multipartite systems. One reason for this is that most questions of information processing do not make sense without some notion of systems being composed of separate subsystems. From these assumptions I derive that systems combine according to a *tensor product rule*. This is of independent interest since it sheds light on where this rule comes from in quantum theory.

From hereon, the notion of a *type* of system is broadened. Thus multipartite systems can come in different types, where a particular type of multipartite system is composed of n_A single systems of type A, n_B single systems of type B, and so on. In all of this section, a system A or B refers to a system of some specific type, that may itself be a composite system.

Begin with the idea that, given a system A, it is possible to identify some operations as *operations on system A alone* and that, in particular, the fiducial measurements for system A are operations on system A alone. (Without this, one might say that we have no business speaking of separate systems in the first place.)

Assumption 4. Local operations commute. Consider a joint system composed of systems A and B. Suppose that an operation is performed on system A alone with outcome o_A and an operation on system B alone with outcome o_B. The final unnormalized state of the joint system does not depend on the order in which the operations were performed. In particular, this implies that the joint probability of getting outcomes o_A and o_B does not depend on the ordering of the operations.

This assumption means that operations can be regarded as performed simultaneously on systems A and B without ambiguity. It also implies

Corollary 1. The no-signaling principle. If an operation is performed on system A, it is not possible to get information about which operation was performed by measuring system B.

The proof of the corollary is straightforward. Suppose that an operation is performed on system A first, followed by an operation on system B. Whichever operation was performed on system A, the marginal probability of outcome o_B is equal to the probability of o_B in the case that the operation on system B came first. The probability of o_B is thus independent of the operation on system A.

Assumption 5. The global state assumption. The global state of a multipartite system can be completely determined by specifying joint probabilities of outcomes for fiducial measurements performed simultaneously on each subsystem.

Note that while the global state assumption is satisfied in quantum theory and in classical probability theory, it need not be satisfied in an arbitrary theory. For example, it is not true in the case of quantum theory defined over a real Hilbert space [30–32]. So this assumption has significant content.[4]

It follows from these two assumptions that the global state of a multipartite system can be written in the form of a vector of joint probabilities. For example, for a bipartite system AB, it will look like this:

$$\vec{P}^{AB} \equiv \left(\begin{array}{c} P(a=1,b=1|X=1,Y=1) \\ P(a=1,b=2|X=1,Y=1) \\ \vdots \\ \hline P(a=1,b=1|X=1,Y=2) \\ P(a=1,b=2|X=1,Y=2) \\ \vdots \\ \hline \vdots \end{array} \right). \tag{14}$$

$P(a{=}i,b{=}j|X{=}k,Y{=}l)$ is the joint probability of getting outcomes i and j when fiducial measurements k and l are performed on the two subsystems. The no-signaling principle implies

$$\sum_j P(a=i,b=j|X=k,Y=l) = \sum_{j'} P(a=i,b=j'|X=k,Y=l') \ \ \forall \ \ i,k,l,l', \tag{15}$$

$$\sum_i P(a=i,b=j|X=k,Y=l) = \sum_{i'} P(a=i',b=j|X=k',Y=l) \ \ \forall \ \ j,k,k',l. \tag{16}$$

The *reduced state* for system A (analogous to the reduced state in quantum theory, or marginal probabilities in classical probability theory) is given by

[4] Arguably, this is not the case for the assumption that local operations commute, which may be regarded as part of the definition of what we mean by an operation being on system A alone. Not wishing to be dogmatic on this point, I have listed this principle with the other assumptions. We should distinguish, however, the implied no-signaling principle from *the impossibility of superluminal signaling*, which is a contingent fact that as far as we know is true in our universe. To see the difference, consider that in the nonrelativistic quantum mechanics of particles, the no-signaling principle is valid, yet superluminal signaling is possible. In the present framework, the impossibility of superluminal signaling would imply an upper bound on the velocity of systems and that Alice cannot carry out an operation on Bob's system if she is spacelike separated from it. But I shall not use such notions, or indeed any notion of spacetime structure.

$$\vec{P}^A = \left(\begin{array}{c} P(a=1|X=1) \\ P(a=2|X=1) \\ \vdots \\ \hline P(a=1|X=2) \\ P(a=2|X=2) \\ \vdots \\ \hline \vdots \end{array} \right), \qquad (17)$$

where

$$P(a=i|X=j) = \sum_{i'} P(a=i,b=i'|X=j,Y=j'). \qquad (18)$$

Here, a and X are the outcome and fiducial measurement for the system whose reduced state is defined, and b and Y are the outcome and fiducial measurement for the other system. The no-signaling conditions of Eqs. (13) and (16) ensure that the sum on the right is independent of the choice of j'.

As seen in the last section, a particular theory specifies a set $\mathcal{S}$ of allowed states for each type of system. This applies also for each type of multipartite system. There is, however, a constraint.

Constraint 2. Suppose that $\vec{P}^{AB} \in \mathcal{S}^{AB}$, where $\mathcal{S}^{AB}$ is the set of allowed states for the joint system. Suppose that $\vec{P}^A$ is the reduced state for system A corresponding to $\vec{P}^{AB}$. Then $\vec{P}^A \in \mathcal{S}^A$, where $\mathcal{S}^A$ is the set of allowed states for system A.

That systems combine according to a tensor product rule is asserted by the following three theorems. Proofs are in Appendix B.

Theorem 1. Denote the vector spaces containing the vectors $\vec{P}^{AB}$, $\vec{P}^A$, and $\vec{P}^B$ by V^{AB}, V^A, and V^B, respectively. Then one can identify

$$V^{AB} = V^A \otimes V^B.$$

Theorem 2. Any $\vec{P}^{AB} \in \mathcal{S}^{AB}$ can be written

$$\vec{P}^{AB} = \sum_i r_i \vec{P}_i^A \otimes \vec{P}_i^B, \qquad (19)$$

with the r_i real, $\vec{P}_i^A \in \mathcal{S}^A$ and $\vec{P}_i^B \in \mathcal{S}^B$. Both $\vec{P}_i^A$ and $\vec{P}_i^B$ can be taken to be normalized and pure.

Theorem 3. Consider a transformation on system A alone defined by

$$\vec{P}^A \to \vec{P}'^A = M^A \cdot \vec{P}^A.$$

The transformation of the joint system is given by

$$\vec{P}^{AB} \to \vec{P}'^{AB} = (M^A \otimes I) \cdot \vec{P}^{AB}.$$

Recall that transformations include probabilistic transformations that decrease the normalization of the state. Thus an immediate corollary of Theorem 3 is

Corollary 2. If a measurement is performed on system A alone, with state $\vec{P}^A$, the probability of a particular outcome is given by

$$\vec{R} \cdot \vec{P}^A = (\vec{R} \otimes \vec{I}) \cdot \vec{P}^{AB}. \qquad (20)$$

Here, $\vec{I}$ is a vector representing the identity measurement, that is $\vec{I} \cdot \vec{P}^B = |\vec{P}^B| \forall \vec{P}^B \in \mathcal{S}^B$. The way things are set up, $\vec{I}$ is not unique but can always be taken to be $(1,\ldots,1|0,\ldots,0|0,\ldots,0|\cdots)$.

Much follows from these theorems and corollary.

Collapsed states. Suppose that an operation is performed on a system A in a state $\vec{P}^A$. Suppose that the operation has outcomes i such that the final normalized state conditioned on outcome i is given by $\vec{P}_i^A \equiv M_i \cdot \vec{P}^A / |M_i \cdot \vec{P}^A|$. The change in the state of system A is analogous to the quantum-mechanical collapse of the state vector. If systems A and B begin in some joint state $\vec{P}^{AB}$, and a measurement is performed on system A, then the final state of system B, conditioned on a particular outcome for the measurement, is also unambiguously determined. Thus this "collapse" is also well-defined "at a distance." Typically, similar questions of interpretation arise in theories in this framework as do in quantum theory. Is this collapse a real process? A change in an agent's degrees of belief following her measurement? And so on.

Entanglement and nonlocality. In Theorem 2, a joint state of a system AB is written as a linear sum of direct product states. Note that the theorem does not assert that a joint state of AB can be written as a convex combination of direct product states. In general, there will be joint states that cannot be written in this form. These are the *entangled* states of the theory. Entanglement is distinct from nonlocality, where the latter means violation of a Bell inequality. Thus (i) there are theories such as classical theories that have no entanglement or nonlocality, (ii) there may be theories that have entanglement but no nonlocality, and (iii) there are theories, such as quantum theory and GNST developed below, that have both entanglement and nonlocality, although these may not coincide.[5]

Multipartite systems. The state of a multipartite system can be written as a vector $\vec{P}^{AB\cdots Z} \in V_A \otimes V_B \otimes \cdots \otimes V_Z$. This vector can be written as a linear sum of direct product states $\Sigma_i r_i \vec{P}_i^A \otimes \vec{P}_i^B \otimes \cdots \otimes \vec{P}_i^Z$, with $r_i \in \mathbb{R}$, $\vec{P}^A \in \mathcal{S}^A$, and so on. A transformation on system A alone takes the form $M \otimes I \otimes \cdots \otimes I$, and similarly for transformations on $B, \ldots, Z$ alone. These extensions of the above theorems follow, since those theorems were stated for arbitrary bipartite systems AB and included the fact that A and B may themselves be composite.

[5] It is clear that entanglement is necessary for nonlocality. But in quantum theory there are entangled mixed states that are local [33,34], hence entanglement is not sufficient for nonlocality. In GNST, on the other hand, entanglement and nonlocality do coincide. This is because if one can write down a local model for a particular state in GNST, then the model will itself define a convex decomposition of that state into product states allowed by the theory. This is not true in quantum theory because arbitrary local models can employ probability assignments not corresponding to any quantum state.

Finally, recall that a theory, in addition to specifying the set $\mathcal{S}$ of allowed states for each type of system, must also specify the set $\mathcal{T}$ of allowed transformations.

Definition 1. A transformation on system A is *well-defined* if $(M_i^A \otimes I)\cdot \vec{P}^{AB} \in \mathcal{S}^{AB}$ whenever $\vec{P}^{AB} \in \mathcal{S}^{AB}$, for all types of system B.

This definition corresponds to the fact that in quantum theory, allowed transformations must be completely positive maps (and not, e.g., merely positive maps). An obvious constraint is the following.

Constraint 3. For each type of system, all transformations $\in \mathcal{T}$ must be well-defined.

A natural assumption is the following.

Assumption 6. If $\vec{P}^A \in \mathcal{S}^A$ and $\vec{P}^B \in \mathcal{S}^B$, then $\vec{P}^A \otimes \vec{P}^B \in \mathcal{S}^{AB}$.

A final assumption that is convenient is the following.

Assumption 7. A theory first specifies a set $\mathcal{S}$ of allowed states for each type of system (including multipartite systems). All transformations that are well-defined are then allowed transformations.

This assumption is indeed satisfied by all the theories considered below, including classical theories, quantum theory, GNST, and GLT. It is nice because it means that a theory is completely specified once the allowed types of system are specified, along with the set $\mathcal{S}$ of allowed states for each type. In this case, Assumption 7 defines the set $\mathcal{T}$. The way things are set up, each of the sets $\mathcal{O}$, $\mathcal{M}$, and $\mathcal{R}$ is in turn defined by $\mathcal{T}$. Assumption 7 also ensures that certain other obvious constraints hold that do not then need to be stated separately. For example, it implies that if $M \in \mathcal{T}$ and $N \in \mathcal{T}$, then $M\cdot N \in \mathcal{T}$. Along with Constraint 3, it implies that if $M^A \in \mathcal{T}^A$, then $M^A \otimes I^B \in \mathcal{T}^{AB}$. Finally, Assumption 7, along with Assumption 6 and Constraint 2, implies that if a procedure consists in introducing an ancilla to system A, performing some joint transformation on A and ancilla and then throwing away the ancilla, then the corresponding transformation on A alone is $\in \mathcal{T}^A$.

The fact that transformations have to be well-defined yields one of the main insights of this work. There is a rich interplay between the set of allowed states, the allowed dynamics, and the information processing possibilities that a theory offers. For example, if a theory is modified by enlarging the set of allowed states (adding supercorrelated states to quantum theory, perhaps), one might naively think that this must increase the information processing possibilities. However, enlarging the set of allowed states may well have the effect of decreasing the set of allowed transformations, in which case the effect may well be the opposite.

III. A BRIEF NOTE ON AMBIGUITIES

There are a couple of points that deserve a mention here in case it be thought that they cause problems (this section may perhaps be omitted on a first reading). First, two theories may be identical in their structure, that is the sets $\mathcal{S}$, $\mathcal{T}$, $\mathcal{O}$, $\mathcal{R}$, and $\mathcal{M}$ of allowed states, transformations, operations, outcomes and measurements, could be mathematically identical in each theory, yet the theories be different physically because the mathematical objects are assigned to different physical objects. For example, a particular preparation device could be associated with one state in one theory and another state in the other theory.

Second, one theory could be made to look different, that is have different sets $\mathcal{S}$, $\mathcal{T}$, $\mathcal{O}$, $\mathcal{R}$, and $\mathcal{M}$, simply because different measurement devices are chosen to correspond to fiducial measurements. Thus in quantum theory the state of a qubit could be specified by the probabilities for the outcomes of spin measurements in the x, y, and z directions. The normalized states in $\mathcal{S}$ are then a ball. Equally, the quantum state could be specified by the probabilities for measurements in the x, y, and n directions, where $\vec{n}=(1/\sqrt{2})(\vec{x}+\vec{z})$. In this case, the normalized states in $\mathcal{S}$ are a nonspherical ellipsoid. A fiducial set may even have different numbers of measurements and outcomes. For example, any quantum state can be expressed by giving the probabilities of the outcomes for a single, informationally complete POV measurement [15]. The important thing is that the outcomes of the fiducial measurements in the new formulation are represented by linearly independent vectors in the old formulation. Thus there is an invertible matrix N such that the two formulations are related by $\vec{P}'=N\cdot\vec{P}$, $\vec{R}'^T=\vec{R}^T\cdot N^{-1}$, and $M'=N\cdot M\cdot N^{-1}$. The theory makes the same predictions since $\vec{R}'\cdot\vec{P}'=\vec{R}\cdot\vec{P}$, and so on.

The first of these points means that in order to compare the predictions of two theories, one has to know which physical devices different preparations and operations correspond to. But being primarily interested in the information processing properties of theories, we can ignore this issue and concentrate on the structure of the theories. The second point ensures that we can do this unambiguously. The structure of a theory and the conclusions drawn for information processing do not depend on which measurements are chosen for the fiducial set.

IV. SOME DIFFERENT THEORIES

It is useful to see examples of theories that can be described in this framework. The most important are classical theories and quantum theory. Two others are GLT and GNST. All of these theories satisfy Assumption 7, which means that each is completely determined by the set $\mathcal{S}$ of allowed states for each type of system.

A. Classical theories

Suppose that for some particular type of system, the fiducial set can be chosen as a single measurement with d outcomes, and that any (possibly subnormalized) probability distribution over these outcomes corresponds to a (possibly subnormalized) allowed state. In this case, the system is *classical*. A classical theory is one for which all systems are classical. The most comprehensive classical theory is the one for which there is a type of system for every $d \geqslant 1$. For a classical system, $\mathcal{S}$ is a simplex. Pure states are represented by vectors $\vec{e}_i$, with a 1 for the ith component and 0s elsewhere. The state of a bipartite system of two classical systems is also represented by a vector from a probability simplex, the entries being the joint probabilities for outcomes i and j when the fiducial measurement is performed on each system. An allowed transformation M must map a pure state $\vec{e}_i$ to another allowed state. It is easy to show that each entry of M must be positive, and the sum of each column must be $\geqslant 0$ and $\leqslant 1$. In the case that M preserves normalization, it is a stochastic matrix.[6] The set $\mathcal{R}$ is a hypercube.

Consider, for example, an ordinary die which can exist in six different deterministic states. The $\vec{P}$ vector is six dimensional and gives the probabilities that the die's uppermost face is 1,2,…,6. An example of a measurement is one that asks, is the uppermost face 1 or 2? The yes outcome corresponds to the vector $\vec{R}=(1,1,0,0,0,0)$. The state of two dice, A and B, can be written as a 36-dimensional vector, whose entries are the probabilities for the uppermost faces being 11,12,…,66.

Suppose that the reduced states of the two dice are given by $\vec{P}^A=\vec{P}^B=\frac{1}{6}(1,1,1,1,1,1)$. One possible joint state compatible with $\vec{P}^A$ and $\vec{P}^B$ is a direct product,

$$\vec{P}^{AB}=\vec{P}^A\otimes\vec{P}^B=\frac{1}{36}\begin{pmatrix}1\\1\\\vdots\\1\end{pmatrix}.$$

This corresponds to the two dice being uncorrelated. But another possible joint state with the same reduced states is

$$\vec{P}^{AB}_{ij}=\begin{cases}1/6 & i=j,\\ 0 & \text{otherwise}.\end{cases}$$

This corresponds to perfect correlation and obviously cannot be written as a direct product. Of course there is no entanglement or nonlocality in this theory.[7]

B. Quantum theory (in finite dimensions)

Quantum theory only allows certain types of system. For example, there are no systems that can be described with two fiducial measurements each with two outcomes. A qubit can be described by three fiducial measurements with two outcomes, e.g., spin measurements in the x, y, and z directions. Once a set of fiducial measurements is chosen quantum theory tells us what the allowed states $\vec{P}$ are. In the simple case of a qubit, the set of normalized states is the Bloch ball. In the case of higher dimensional quantum systems it does not appear to be so easily characterized (except via the usual quantum formalism of course). The transformations that are well-defined, in the sense of Definition 1, correspond precisely to the trace-nonincreasing completely positive maps. It is usually assumed that any such map corresponds to a physically possible operation, thus Assumption 7 is satisfied. Any set of $\vec{R}_i$ with $0\leqslant\vec{R}_i\cdot\vec{P}\leqslant 1\,\forall\; i\,\forall\;\vec{P}\in\mathcal{S}$ and $\Sigma_i\vec{R}_i\cdot\vec{P}=1\,\forall\;\vec{P}\in\mathcal{S}$ is a positive operator-valued measurement in the usual formalism.

There is nothing new in the fact that quantum states can be represented as real vectors and transformations as matrices acting on these vectors. It is well known that Hermitian operators in d dimensions form a d^2-dimensional real vector space, with an inner product given by $\mathrm{Tr}(AB)$. Linear completely positive maps correspond to $d^2\times d^2$ matrices acting on this space. But the present framework does not correspond exactly to this representation (e.g., it is possible that $\vec{P}\cdot\vec{P}>1$), so it is useful to see an example. A qubit whose state is spin up in the z direction can be written

$$\vec{P}=\left(\begin{array}{c}P(\uparrow|x)\\P(\downarrow|x)\\\hline P(\uparrow|y)\\P(\downarrow|y)\\\hline P(\uparrow|z)\\P(\downarrow|z)\end{array}\right)=\left(\begin{array}{c}1/2\\1/2\\\hline 1/2\\1/2\\\hline 1\\0\end{array}\right),$$

where $P(\uparrow|x)$ is the probability of obtaining spin up when measuring in the x direction, and so on. It can now be verified that if, for example, spin is measured in the n direction, where $\vec{n}=(1/\sqrt{2})(\vec{x}+\vec{z})$, then the up outcome corresponds to the vector

$$\vec{R}=\left(\frac{1}{2\sqrt{2}},\frac{-1}{2\sqrt{2}}\,\middle|\,\frac{1}{2},\frac{1}{2}\,\middle|\,\frac{1}{2\sqrt{2}},\frac{-1}{2\sqrt{2}}\right).$$

This vector is not unique. Any vector $\vec{R}'=\vec{R}+\vec{C}$, where $\vec{C}\cdot\vec{P}=0\,\forall\;\vec{P}\in\mathcal{S}$, represents the same measurement outcome. The unitary transformation usually written as the Pauli matrix σ_z would correspond to

$$M=\left(\begin{array}{cc|cc|cc}0&1&0&0&0&0\\1&0&0&0&0&0\\\hline 0&0&0&1&0&0\\0&0&1&0&0&0\\\hline 0&0&0&0&1&0\\0&0&0&0&0&1\end{array}\right).$$

C. Generalized nonsignaling theory

Suppose that for any pair $n,k>1$, there is a corresponding type of single system, whose state can be described by a set

[6] In this work, a stochastic matrix is a not necessarily square matrix, with non-negative entries, whose columns each sum to 1.

[7] There is nothing difficult in the preceding remarks. But part of the aim of Sect. II B is to deflate the significance of the tensor product rule for combining systems in quantum theory. Thus it is useful to note that a similar rule arises quite naturally in what is essentially classical probability theory. The quantum tensor product rule does not have to be regarded, as it frequently is, as a mysterious replacement for the Cartesian product used in combining deterministic classical states. If quantum states (even pure ones) are more analogous to probabilistic classical states than anything else—in other words if some version of the *epistemic interpretation* of the quantum state is correct—then a tensor product rule is exactly what one would expect. Thus one way of viewing the tensor product is as evidence for the epistemic interpretation.

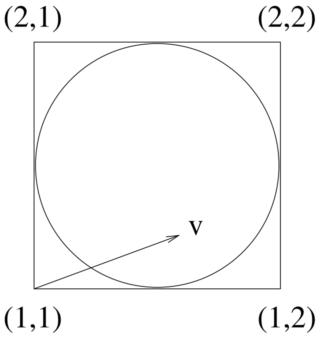

FIG. 1. The space of normalized states for a gbit in GNST corresponds to the square. If the measurements $X=1$ and $X=2$ are associated with spin measurements in the z and x directions, then the space of states for a quantum-mechanical qubit corresponds to the circle.

of n fiducial measurements, each with k outcomes. Call this an (n,k) system.[8] For a single system, allow any state $\vec{P}$, provided the entries of $\vec{P}$ are between 0 and 1 and Eq. (3) is satisfied. For multipartite systems, allow any state $\vec{P}$, provided entries are between 0 and 1, Eq. (3) is satisfied, and the no-signaling conditions of Eqs. (15) and (16) are satisfied for all bipartite splittings. The resulting theory is generalized nonsignaling theory.

It is useful to see some examples of systems in this theory. The simplest kind of single system has two binary fiducial measurements. This type of system plays a role somewhat analogous to that of a classical bit or a qubit, so from hereon it is called a *gbit* (for generalized bit). The space of possible normalized states is shown in Fig. 1. There are four pure states, which correspond to the four ways of assigning definite outcomes to the $X=1$ and $X=2$ fiducial measurements. In the figure, these are represented by (1,1), (1,2), (2,1), and (2,2) where (1,2), for example, is the state which returns $a=1$ for the $X=1$ measurement and $a=2$ for the $X=2$ measurement, and is also represented by $\vec{P}=(1,0|0,1)$. Thus pure states of single systems have a definite outcome for each fiducial measurement—there is no uncertainty principle. As noted in the figure, if the measurements $X=1$ and $X=2$ are associated with spin measurements in the z and x directions, then we can include possible states of a qubit in the diagram, and these form a circle inscribed in the square. Qubits of course have an extra degree of freedom, namely spin in the y direction. For (3,2) systems the space of states is a cube, with an inscribed sphere (the Bloch sphere) representing quantum states.

Consider the possible transformations of a gbit (for simplicity, restrict attention to those that preserve normalization). An allowed transformation will transform the square in such a manner that all points remain in the square, otherwise the transformation is not well-defined in the sense of Definition 1. The transformations of Figs. 2 and 3 are allowed. But the transformation of Fig. 4 is not allowed. Transformations in quantum theory are less restricted because the requirement is only that points in the circle are transformed into points in the circle. So a rotation of $\pi/4$, as in Fig. 4, is fine, and indeed corresponds to the well-known $\pi/8$ gate.

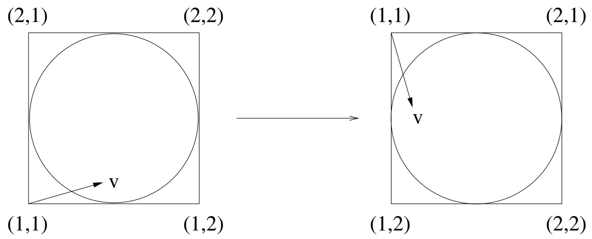

FIG. 2. An allowed transformation.

It begins to look as if the dynamics of single systems in GNST is rather simple. Indeed, this is the case. Section VI contains a theorem that states that for single systems in GNST, allowed transformations correspond essentially to relabelings of measurements and outcomes, and probabilistic combinations thereof. Thus in a sense, the dynamics is classical. Despite this, the dynamics does contain possibilities that quantum dynamics does not. Consider a (3,2) system, whose space of normalized states is a cube, with the quantum Bloch sphere inscribed. A possible transformation is a reflection in the center of the sphere. This corresponds to the universal NOT gate of quantum theory, which is not an allowed transformation since it is not completely positive.

The multipartite states of GNST are noteworthy in that they include states that are more nonlocal than quantum theory allows. For example, given a bipartite system of two gbits, the following is a possible state.

$$XY = \left.\begin{matrix} 11 \\ 12 \\ 21 \end{matrix}\right\} \rightarrow P(a=1,b=1|XY) = P(a=2,b=2|XY) = 1/2, \tag{21}$$

$$XY = 22 \rightarrow P(a=1,b=2|XY) = P(a=2,b=1|XY) = 1/2. \tag{22}$$

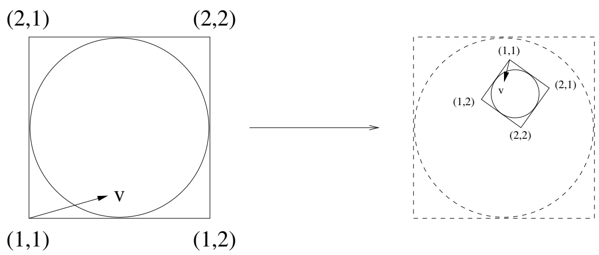

FIG. 3. Another allowed transformation.

[8] A more general theory would include further types of system with different numbers of outcomes for different fiducial measurements. I ignore this possibility. I do not believe that it would change much beyond introducing uninteresting complications into some of the proofs.

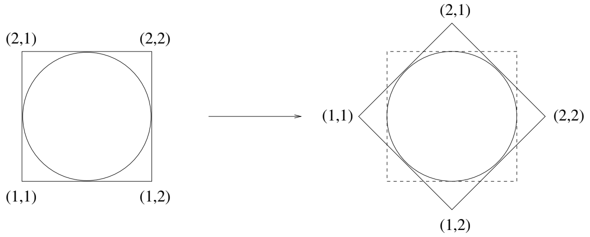

FIG. 4. A forbidden transformation.

The correlations obtained from fiducial measurements on this state return a value of 4 for the left-hand side of the following inequality:

$$P(a=b|11)+P(a=b|12)+P(a=b|21)+P(a\neq b|22)\leqslant 3 \tag{23}$$

(this is the CHSH inequality [35] written in a slightly different form than usual). These correlations cannot be obtained from measurements on any quantum state since by Tsirelson's theorem [36], quantum states can only reach a maximum of $2+\sqrt{2}$.[9]

Information processing in GNST is discussed in Sec. VII. The theory's permissiveness with respect to states implies that some things can be achieved that are impossible in quantum theory. These include 1-2 oblivious transfer, van Dam's scheme for the easy solution of communication complexity problems [25], and a kind of superquantum memory. The restricted nature of the dynamics, however, implies that there is no teleportation or super-dense coding. The theorems of Sec. VI give evidence that computation is no better than classical.

D. Generalized local theory

Suppose that, as in GNST, for any pair $n,k>1$, there is a corresponding type of system, whose state can be defined with n fiducial measurements with k outcomes. As in GNST, all $\vec{P}$ with entries between 0 and 1 satisfying Eq. (3) are allowed states. The only multipartite states allowed, however, are those for which the fiducial measurements return local (non-Bell-violating) correlations. This defines generalized local theory.

As in GNST, the pure states of single systems in this theory are those that have a deterministic outcome for each fiducial measurement. Since multipartite states are local with respect to fiducial measurements, the pure states of a multipartite system are precisely those in which each subsystem is in a deterministic pure state. An arbitrary state of a multipartite system is a convex mixture of these. It follows that no state in this theory can violate a Bell inequality, even if nonfiducial measurements are performed. Hence the name.

GLT is more general than quantum theory in allowing arbitrary single system states, but more restricted in not allowing nonlocal states. As described in Sec. VII, GLT allows 1–2 oblivious transfer. Computation in GLT, however, is efficiently simulable by a classical computer.

E. Other possibilities

There are other possibilities that would be interesting to investigate. For example,

(i) A theory that is essentially quantum theory but with only separable states allowed.

(ii) A theory in which the state of a single system must be a quantum state, but in which the state of a multipartite system can be anything, as long as the no-signaling principle and the restriction that the reduced states for the individual subsystems must be quantum are satisfied. The latter idea has been investigated in Ref. [37], where it is shown, amongst other things, that Tsirelson's theorem still holds.

V. GENERIC PROPERTIES OF THEORIES

One of the reasons for introducing a framework encompassing many different theories is that it is interesting to identify properties of theories that are generic, in the sense that they are shared by all or most theories in the framework. Some features, usually thought of as specifically quantum, are present in all theories in our framework except theories that are classical (in the sense of Sec. IV A). Thus classical theories are very special! These features include the fact that mixed states do not always have a unique decomposition into pure states, and a no-go theorem for universal cloning. More exact statements of these claims are given in this section. Proofs are in Appendix C. It is tedious to write always *all theories in the framework*, so from hereon this is shortened to *all theories*, taking the assumptions of Sec. II as read.

First, consider the following.

Theorem 4. Suppose that for each type of system in some theory, every mixed state has a unique decomposition into pure states and $\vec{0}$. Then the theory is classical.

The next theorem concerns the disturbance of systems on measurement and is due in part to Howard Barnum, Matthew Leifer, and Alexander Wilce [38]. Say that a transformation *disturbs* a state $\vec{P}$ if there is no constant c such that $M\cdot\vec{P}=c\vec{P}$. This means that, conditioning on the outcome corresponding to this transformation, the state is no longer $\vec{P}$. A transformation is nondisturbing if no pure state is disturbed and an operation $\{M_i\}$ is nondisturbing if all M_i are nondisturbing.

Theorem 5. For any system, let V be the vector space in which states are defined, and let V_S be the subspace spanned by $\mathcal{S}$. Then V_S can be written as a direct sum, $V_S=\oplus_i V_i$, where the V_i are subspaces of V_S, such that

(i) Every pure state $\vec{P}$ is contained in some V_i.

[9] These nonsignaling superquantum correlations were written down by Khalfi and Tsirelson, [16] and were independently introduced by Popescu and Rohrlich [17]. Other examples of superquantum correlations, involving more measurements or parties, are given in Ref. [18]. The latter are also allowed in GNST.

(ii) A nondisturbing transformation is of the form $M = \oplus_i e_i I_i$, where $0 \leqslant e_i \leqslant 1$, and I_i is the identity on V_i.

It follows that nondisturbing operations have the same outcome probabilities for pure states in the same V_i, and thus cannot distinguish them. For a classical system, each V_i contains exactly one pure state. For a quantum system without superselection rules, V_S cannot be further decomposed into a direct sum. Nondisturbing operations have the same outcome probabilities for all pure states, each transformation being proportional to the identity on V_S. An example of such an operation would be to toss a coin, without interacting with the system at all, and to output the result. For a quantum system with superselection rules, pure states from the same sector are elements of the same V_i, and different sectors correspond to different V_i.

Theorem 5 has implications for cloning. Cloning refers to the following procedure:

(i) Begin with a system A in a pure state. Denote its state $\vec{P}$.

(ii) Introduce a system B of the same type, prepared in a standard state $\vec{Q}$. The state of the joint system is $\vec{P} \otimes \vec{Q}$.

(iii) A joint transformation M acts on the pair of systems such that the final state is $M \cdot (\vec{P} \otimes \vec{Q}) \propto \vec{P} \otimes \vec{P}$.

A *deterministic universal cloning procedure* always succeeds and works on all pure states. It implies the existence of a normalization-preserving M and a state $\vec{Q}$ such that $M \cdot (\vec{P} \otimes \vec{Q}) = \vec{P} \otimes \vec{P}$ for all pure $\vec{P}$. A *probabilistic universal cloning procedure* is allowed to output a fail outcome, but conditioned on success, the final state must be $\vec{P} \otimes \vec{P}$. There must be a nonzero probability of success for all pure states $\vec{P}$. This type of cloning implies the existence of a nonzero M such that $M \cdot (\vec{P} \otimes \vec{Q}) = c\vec{P} \otimes \vec{P}$ for all pure $\vec{P}$, where c can vary with $\vec{P}$, and $0 < c \leqslant 1$.

Theorem 6. Suppose that in some theory, there is a probabilistic universal cloning procedure for each type of system. Then the theory is classical.

This of course implies that if every system has a deterministic universal cloning procedure, then a theory is classical.

Theorems 4, 5, and 6 apply even to classical theories if extended to mixed states. Thus there are mixed states with a nonunique decomposition into mixed states. All transformations disturb at least one mixed state unless they are proportional to the identity.[10] And cloning of classical mixed states is impossible.[11] One possible interpretation of these remarks is as further evidence that quantum pure states are more akin to classical mixed states than classical pure states.

There are many other questions concerning properties that are common to all theories, or all except classical theories. In Ref. [39], the quantum no-broadcasting theorem is generalized to arbitrary nonclassical theories within a framework closely related to the present one. It can also be shown that all theories in that framework have an infinite de Finetti theorem, [40] and that polynomially sized computations in any of these theories can be simulated classically in polynomial space [41]. Features such as these can be regarded as arising solely from the assumptions that were made in setting up the framework.

VI. DYNAMICS IN GNST AND GLT

Part of the motivation of this work is to consider which features of a theory, in particular those features related to information processing, arise from which assumptions. It is particularly interesting if significant features, such as the no-cloning theorem above, arise from very minimal assumptions and are thus shared by a broad class of theories. Another part of the motivation is to investigate theories that are different from those we already know about. These theories need not even be empirically adequate; a compare and contrast exercise will still be useful to learn more about those theories that *are* empirically adequate. Thus the next two sections are devoted to a detailed investigation of GNST and GLT.

In Sec. IV C the dynamics of a gbit was briefly discussed. There are four pure states of a gbit, corresponding to the four ways of assigning definite outcomes to the two measurements. The space of normalized states is a square, with a normalization-preserving transformation being a linear transformation of this square. Let us consider more general types of system in GNST and GLT, but continue to focus on normalized systems and normalization-preserving transformations, i.e., operations corresponding to a single matrix, $\{M\}$. For this section and the next, *transformation* means *normalization-preserving transformation*, with the investigation of probabilistic transformations left for future work.

The space of normalized states of an (n,k) system is a polytope, the vertices corresponding to pure states. Pure states are of the form

[10] This is not at all surprising if put into more prosaic terms. Consider that a die is in a state such that the probability of each face being uppermost is $\frac{1}{6}$. Suppose that the die is measured, to find out which face is uppermost, and the value 1 found. Then, if it is assumed that the measurement operation was done in the most obvious way, the state after measurement is not $\frac{1}{6}(1,1,1,1,1,1)$, but (1,0,0,0,0,0). Of course the measurement operation may be such that the die is recast after the outcome is obtained, resulting in a final state of $\frac{1}{6}(1,1,1,1,1,1)$. But then an initial state of (1,0,0,0,0,0) would be disturbed.

[11] Suppose that Alice prepares a die in one of two ways, each corresponding to a probability distribution over the different faces. The first prepares, say, the state $\frac{1}{12}(6,2,1,1,1,1)$ and the second the state $\frac{1}{6}(1,1,1,1,1,1)$. The die is given to Bob who is required to perform a cloning operation. This means that Bob must prepare another die such that if Alice used the first preparation, its state is $\frac{1}{12}(6,2,1,1,1,1)$, and if she used the second, then $\frac{1}{6}(1,1,1,1,1,1)$. Furthermore, if the dice are measured after Bob's operation, the results *must not be correlated*. This last clause prevents Bob from using a device that simply reads the uppermost face of the die and prepares another in the same state. It is easy to see that even if Bob's cloning procedure is allowed to be probabilistic, he cannot do it.

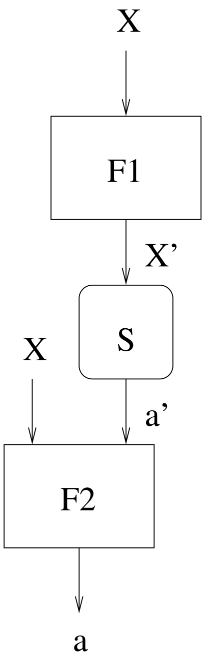

FIG. 5. Transformations of single systems in GNST and GLT can always be represented as the appending of classical circuits as shown here, or as convex combinations of transformations of this type. If a fiducial measurement X is performed on the transformed system, this can be thought of as performing fiducial measurement X' on the original system, where $X'=F1(X)$ for some function $F1$. When measurement X' is performed on the original system, outcome a' is obtained with some probability. The probability of obtaining outcome a for the measurement X on the transformed system is equal to the probability of obtaining an outcome a' such that $a=F2(X,a')$, for some function $F2$.

$$\vec{P}=(0\cdots 1\cdots 0|0\cdots 1\cdots 0|\cdots).$$

Allowed transformations must take points in the polytope to points in the polytope. This condition is so restrictive that the following theorem holds.

Theorem 7. Normalization-preserving transformations of single systems in GNST or GLT, thought of as active, correspond to passive transformations that simply relabel fiducial measurements and outcomes, or to convex combinations of such. Equivalently, for a transformation of an (n,k) system, the matrix M representing the transformation can be written

$$M=\left(\begin{array}{c|c|c} M_{11} & \cdots & M_{1n} \\ \hline \vdots & & \vdots \\ \hline M_{n1} & \cdots & M_{nn} \end{array}\right),$$

where M_{ij} is a $k\times k$ matrix, and where $M_{ij}=\alpha_{ij}S_{ij}$, for S_{ij} a stochastic matrix, $0\leqslant\alpha_{ij}\leqslant 1$, and $\Sigma_j\alpha_{ij}=1$.

A useful pictorial representation of this theorem is given in Fig. 5. A related result is as follows.

Theorem 8. The only measurements on single systems in GNST or GLT are fiducial measurements, possibly with outcomes relabeled, or correspond to convex combinations of such.

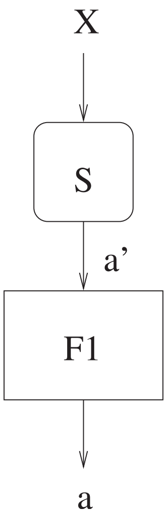

FIG. 6. Measurements on single systems in GNST or GLT can always be performed via a procedure of the type illustrated here, or via a convex combination of such procedures. First, fiducial measurement X is performed and outcome a' is obtained. The outcome a of the complete measurement is then $a=F1(a')$. This result applies to measurements with an arbitrary number of outcomes.

Theorem 8 is illustrated pictorially in Fig. 6. The proofs of Theorems 7 and 8 are contained in Appendix D.

In the case of GNST, the following theorem holds for a bipartite system of two gbits, and suffices to characterize the normalization-preserving transformations of such a system.

Theorem 9. Consider a system of two gbits in GNST, and suppose that a normalization-preserving transformation is performed. Suppose that this transformation is followed by the fiducial measurements X,Y on the two subsystems, with outcomes a,b. The joint probability of obtaining outcomes a,b is equal to that obtained from a convex combination of procedures of the following kind. First, perform a fiducial measurement X' on one of the gbits, where X' may depend on X and Y. Denote the outcome a'. Then perform a fiducial measurement Y' on the other gbit, where Y' may depend on X,Y and on a'. Denote the outcome b'. The final outcome pair (a,b) is a function of X, Y, a', and b'.

Of course this theorem can also be expressed in terms of a formal constraint on the transformation matrix M, but in this case it is more complicated and less enlightening. Theorem 9 may also be understood pictorially, as in Fig. 7.

Theorem 10. In GNST, the only measurements on bipartite systems comprised of two gbits correspond to convex combinations of procedures of the following kind. First, perform a fiducial measurement X on one of the gbits, obtaining an outcome a'. Then perform a fiducial measurement Y on the other gbit, where Y may be a function of a', obtaining an outcome b'. The final outcome is a function of a' and b'.

Theorem 10 is illustrated in Fig. 8. The proof of Theorem 9 is given in Appendix D, with the proof of Theorem 10 following in a similar manner to that of Theorem 8. It is an open question whether similar theorems hold for transformations and measurements on arbitrary multipartite systems in GNST. It can be shown that in GLT, there definitely do exist possibilities for measurements and transformations on multi-

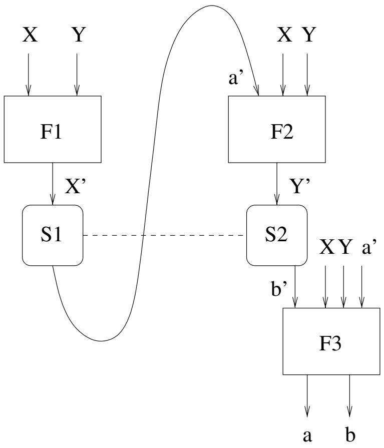

FIG. 7. In GNST, transformations on bipartite systems comprised of two gbits can always be represented by the appending of classical circuits as shown here, or by a similar construction inverted with respect to the two subsystems, or by a convex combination of such. For the construction shown here, this means that if fiducial measurements X,Y are performed on the transformed system, one may think of this as first performing a fiducial measurement X' on one half of the original system, where $X'=F1(X,Y)$. This gives an outcome a'. Then, perform a fiducial measurement Y' on the other subsystem, where $Y'=F2(X,Y,a')$. The final outcome pair (a,b) is determined by a function $F3$ of X, Y, a' and b'.

partite systems that do not reduce to one of the forms presented in this section.

The proofs in Appendix D also make clear the following. The most fine-grained measurements on single systems in GNST or GLT can be represented by a set of vectors $\vec{R}_i$, such that each $\vec{R}_i$ has one element between 0 and 1 and the rest 0. Such an $\vec{R}_i$ is analogous to an effect in quantum theory that is proportional to a one-dimensional projector. A set of $\vec{R}_i$ is analogous to a nondegenerate projective measurement if each $\vec{R}_i$ is a basis vector (one element 1 and the rest 0) and $\Sigma_i \vec{R}_i \cdot \vec{P} = 1 \, \forall \, \vec{P} \in \mathcal{S}$. The corresponding measurement is simply a fiducial measurement, with an $\vec{R}_i$ for each outcome. It is then immediate that, at least with respect to these measurements, there is no Kochen-Specker theorem for single systems in GNST or GLT. Not only is it possible to assign definite outcomes to these measurements in a noncontextual fashion, but each such assignment is in fact an allowed state of the theory. Nonetheless, both GNST and GLT exhibit a different kind of contextuality, introduced by Spekkens [42] and termed *preparation contextuality*. Readers are referred to Ref. [42] for discussion of preparation contextuality. Given the definition, the proofs for GNST and GLT are obvious.

VII. INFORMATION PROCESSING

Using the results obtained for dynamics in GNST and GLT, the information processing possibilities of each theory can be investigated. Rather than attempt something like a general theory of information, this section contains remarks concerning some obvious tasks. Note that there has already been some work investigating the information processing properties of PR boxes, considered merely as abstract correlations. van Dam has shown that they are very powerful for communication complexity problems [25], and this result has recently been extended to noisy PR boxes in Ref. [26]. Others have claimed to show how to do oblivious transfer [19] and bit commitment [21] using PR boxes. However, as pointed out in Ref. [20], the fact that these latter works consider PR boxes only as abstract correlations means that they make assumptions that may not hold in any theory that allows PR boxes.[12] In general, a theory with well defined dynamics is needed before cryptography, or indeed other types of information processing, such as computation, can be discussed. GNST is such a theory.

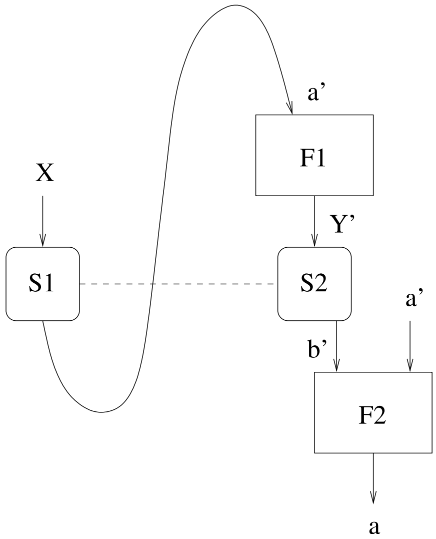

FIG. 8. In GNST, measurements on bipartite systems of two gbits can always be carried out by a procedure like that illustrated here, by a similar procedure inverted with respect to the two subsystems, or by a convex combination of such. For the procedure shown here, this means that first, a fiducial measurement X is performed on one subsystem, and outcome a' is obtained. Then fiducial measurement Y' is performed on the other subsystem, where $Y'=F1(a')$, and outcome b' is obtained. The outcome a of the complete measurement is given by $a=F2(a',b')$. This result applies to measurements with an arbitrary number of outcomes.

The first results concern teleportation and superdense coding (the quantum versions of these tasks were introduced in Refs. [43,44]). The natural analog of a quantum-mechanical singlet in the GNST is a state which, when fiducial measurements are performed, produces the PR box correlations:

$$XY = \left.\begin{matrix} 11 \\ 12 \\ 21 \end{matrix}\right\} \rightarrow P(a=1,b=1|XY) = P(a=2,b=2|XY) = 1/2,$$

[12] One such assumption that as far as I know has not been pointed out is that the shared boxes are trusted to behave like PR boxes by both parties. But one may reasonably ask where did they come from? By whom were they distributed?

$$XY = 22 \rightarrow P(a=1,b=2|XY) = P(a=2,b=1|XY) = 1/2.$$

It can be shown that these correlations represent a pure state—that is a vertex of the polytope of normalized states for two gbits. Further, all vertices of this polytope are either local deterministic correlations (product pure states) or are equivalent to the PR box under local transformations [18].

Theorem 11. It is impossible to teleport an unknown gbit using a single shared PR box.

Proof. This follows easily from Theorem 10. In order to teleport an unknown gbit, Alice must perform some operation or sequence of operations on the gbit and her half of the shared PR box. Without loss of generality, whatever she does may be represented as a single joint measurement, with m outcomes, on the two subsystems. But this measurement can be represented as a convex combination of procedures like that of Fig. 8. Such a procedure will always begin, either by measuring $X=1$ or $X=2$ on the gbit, or by measuring the PR box. In the former case, no information is gained about the value for the other measurement on the gbit and teleportation cannot possibly succeed on all pure states. In the latter case, the shared PR box collapses into a product state which cannot achieve teleportation. ■

Theorem 12. A single shared PR box cannot be used for superdense coding.

Proof. This follows from Theorems 7 and 10. Superdense coding would require that there are four different operations that Alice can perform on her gbit such that, when it is sent to Bob, he can determine unambiguously which was performed by a joint measurement on the two gbits now in his possession. It is easy to see that this is not possible. ■

A. Cryptography

Theorem 13. In GNST, key distribution is possible.

Proof. Key distribution can be achieved in GNST using an Ekert-style protocol [45], in which Alice and Bob first share n pairs of gbits, with each pair in the PR box state. They then test some of their shared systems, to make sure that they really are PR box states, i.e., that they have not been disturbed en route by an eavesdropper. Finally, they measure each remaining gbit pair, using the fiducial measurements $X=1$ and $Y=1$. Assuming that they share perfect PR box states, their measurement outcomes will be perfectly correlated and can be used as a secret key. This protocol is secure because PR box states have a property of being monogamous, much as the entanglement of a singlet is monogamous in quantum theory. Thus consider a tripartite system shared between Alice, Bob, and Eve. If Alice's and Bob's reduced state is the PR box state $\vec{P}_{PR}^{AB}$, then the global state must be of the form $\vec{P}_{PR}^{AB} \otimes \vec{P}^{E}$. The outcome of any measurement performed by Eve is uncorrelated with Alice's and Bob's outcomes. The fact that the PR box correlations are monogamous was shown in Ref. [18]. ■

Recall Theorem 5, which implies that except in classical theories, there are some types of system with pure states (lying in the same subspace V_i) that cannot be distinguished by nondisturbing operations. This motivates the following.

Conjecture 1. In any nonclassical theory, secure key distribution is possible.

The idea is that in any theory with two states that are indistinguishable by nondisturbing operations, it may be possible to find a secure prepare-and-measure protocol [46] that uses those states. Finally, we have the following.

Theorem 14. 1-2 oblivious transfer can be implemented securely in both GNST and GLT.

Proof. In a 1-2 oblivious transfer (introduced in Ref. [47]), Alice must submit 2 bits to Bob in such a manner that Bob can choose to learn either one of the bits or the other, but not both. There is also a security requirement against Alice, who must not be able to learn which of the bits Bob chose. That this task is impossible to implement securely in quantum theory is shown in Ref. [48]. To implement this task in GNST or GLT, Alice sends a gbit to Bob, in a pure state, with the two bits encoded in the outcomes for the $X=1$ measurement and the $X=2$ measurement. Theorem 8 ensures that any strategy employed by Bob is equivalent to his measuring either $X=1$ or $X=2$, or to measuring $X=1$ with some probability p and $X=2$ with probability $1-p$. Thus the protocol is secure against Bob. That it is secure against Alice follows from the fact that, by the no-signaling principle, she cannot determine which measurement Bob did. ■

In classical cryptography, it is known that 1-2 oblivious transfer is equivalent to oblivious transfer [49], and that either can be used to implement arbitrary secure distributed computation [50]. In particular, either can be used to implement bit commitment, hence coin tossing. However, one cannot assume that the standard reductions of classical cryptography hold in a different theory such as GLT, GNST, or quantum theory. Thus it is open whether other two-party cryptographic tasks, such as oblivious transfer, bit commitment or coin tossing, can be implemented securely in GNST or GLT.

B. Computation

For any of the theories in the framework, a natural model of computation may be defined, based on the classical and quantum circuit models. I introduce this model only informally. A particular circuit is assumed to act on n systems, each of the same type, initially prepared in a product state corresponding to the problem input. Instead of k-bit or k-qubit gates, there are transformations that act jointly on k systems. At the end of the computation, the fiducial measurement $X=1$ is performed on each system in order to obtain the output. For a particular theory, it may not be the case that bipartite and single system transformations together are universal, as they are in classical and quantum theory. Thus transformations that act jointly on k systems for any $k>2$ are allowed. But for any circuit family C_n, there must exist some finite k such that all transformations act on k systems or fewer. In addition, it may not be the case that any particular type of system (such as a gbit) is universal for computation in a given theory. So one should keep in mind that circuits may act on other types of system. Finally, in order to define a notion of polynomial time, say, the usual caveats must be assumed. For example, it should be possible for a classical Turing machine to output a description of the ith circuit in time polynomial in i.

Theorem 15. In GLT, any computation can be simulated efficiently by a probabilistic classical computer.

Proof. In GLT, any allowed state of n systems can be written as a convex combination of local deterministic, or pure product, states, in which each system has a definite outcome for each fiducial measurement. A classical simulation of the GLT computation works by storing, at any given time, a local deterministic state of the n systems. This requires an amount of memory linear in n, rather than the exponential amount needed to store a complete description of an arbitrary convex combination. An allowed transformation T, acting on k systems, must take local deterministic states of the k systems to other allowed states of GLT, which in turn are convex combinations of local deterministic states:

$$T(\vec{P}^{LD}) = \sum_i p_i \vec{P}_i^{LD},$$

where the superscript LD indicates a local deterministic state. The classical computer simulating the GLT computation simply updates the stored state $\vec{P}^{LD}$ to $\vec{P}_i^{LD}$ with probability p_i. When the final $X=1$ measurements are performed, the stored local deterministic state will determine the classical computer's output. ■

The computational power of GNST is at present unclear. But it is known that it is very powerful for communication complexity problems.

Theorem 16. In GNST, bipartite communication complexity problems require only constant communication, provided the parties share sufficient PR boxes.

Recall that in a bipartite communication complexity scenario, two separated parties each receive an input, and their task is to compute some joint function of their inputs. Their goal is to minimize the amount of communication. van Dam has shown that if the two parties have a supply of shared PR boxes, then any communication complexity problem can be solved with only constant communication [25]. This result has recently been strengthened: it continues to hold even if the shared PR boxes are noisy, provided the amount of noise is not too great [26]. Contrast the situation in quantum theory, where the inner product problem is known to require n bits of communication to be solved exactly, even with unlimited shared singlets [51].

Finally, we have the following.

Theorem 17. Superquantum memory. In GNST, it is possible to store a 2^n-bit string in only n gbits. Although the whole string cannot be recovered, it is possible to recover the ith bit without error.

Proof. Suppose that the ith bit of the 2^n-bit string we wish to store is given by $f(i_1,\ldots,i_n)\in\{0,1\}$, where $i_1\cdots i_n$ is the binary representation of i. Let $X_1,\ldots,X_n\in\{0,1\}$ be fiducial measurements on the n gbits and $a_1,\ldots,a_n\in\{0,1\}$ the outcomes. (It is easier for this proof to let X_j and a_j take values in {0,1} instead of in {1,2} as elsewhere.) To store the string, prepare a state of n gbits such that

$$P(a_1,\ldots,a_n|X_1,\ldots,X_n) = \begin{cases} 1/2^{n-1} & a_1\oplus\cdots\oplus a_n = f(X_1,\ldots,X_n) \\ 0 & \text{otherwise} \end{cases}, \qquad (24)$$

where $\oplus$ represents addition mod 2. In order to recover the ith bit of the stored string, simply perform the measurement $X_j=i_j$ on each gbit and sum the outcomes mod 2. One may check that the state of Eq. (24) is an allowed state, since it is normalized and nonsignaling. Note that it is indeed impossible to store a 2^n-bit string in only n qubits such that any bit may be recovered. Bounds on quantum memory are derived in Ref. [52]. ■

VIII. DISCUSSION

A. Framework

The framework introduced allows investigation of theories different from either quantum or classical theories. The general idea is that quantum theory can be better understood by viewing it in a context of different possibilities. More specific motivations include the following:

(i) to understand the links between general physical principles and information processing;

(ii) to stimulate the study of computation in models that are more general than quantum theory;

(iii) to address Popescu's and Rohrlich's question of why quantum theory does not allow the PR box correlations;

(iv) to shed light on the interpretive problems of quantum theory by viewing those in a more general context;

(v) to stimulate research into axioms for quantum theory.

As regards single systems the framework is very general indeed. It should be emphasized in particular that linearity of transformations is not assumed, but is derived from the fact that the vector $\vec{P}$ is by definition a complete description of the system.[13] The most important requirements are that local operations commute (Assumption 4), and the global state assumption (Assumption 5), both involving the manner in which separate systems combine to make joint systems. These imply a tensor product rule.

One of the interesting things to emerge from the framework is that certain features, usually thought of as specifically quantum, are possessed by all theories except classical theories. These include the nonunique decomposition of mixed states into pure states, the existence of sets of pure states that cannot be distinguished with nondisturbing operations, and the impossibility of even probabilistic universal cloning. Thus rather than regard quantum theory as special for having these features, a better attitude may be to regard classical theories as special for not having them.

[13] So what of nonlinear modifications of quantum mechanics? These modifications are nonlinear in the sense that they involve a nonlinear Schrödinger equation. In this case, the usual density matrix is no longer a complete description of a quantum system, since the evolution of a system will in general depend not only on the density matrix, but on the particular decomposition into pure states (assuming a proper mixture). If the description of the state is expanded until it is complete, then the action of the dynamics on this new expanded state description will be linear. But such a theory will in general violate one or more of the other assumptions. A list of references on nonlinear quantum theories is given in Ref. [53], and computation in this context is considered in Ref. [3].

How reasonable are Assumptions 4 and 5? Commutativity of local operations is arguably part of what it means to talk about separate systems. In a theory where it fails, any measurement or transformation is essentially a measurement or transformation on all systems at once. It is no longer obvious how to define a reasonable model of computation—how should resources be counted? The case for assuming the commutativity of local operations is also strengthened by the fact that in a spacetime framework, it can be independently motivated by special relativity. It is slightly more difficult to regard the global state assumption as independently compelling. Thus an interesting direction in which to extend this work would be to generalize the framework further by dropping this assumption.

B. Tensor product rule

It is interesting to compare the derivation of the tensor product rule with that of Fuchs [15]. Without going into too much detail, Fuchs assumes that local measurements on two separate systems, A and B, are represented by positive operator-valued measures on Hilbert spaces H_A and H_B. He derives a Gleason-like theorem [54–56] which states that the joint state of the two systems can be represented by an operator on the tensor product Hilbert space $H_A \otimes H_B$, with joint probabilities for outcomes of local measurements given by the standard trace rule.

As Fuchs acknowledges, the proof does not establish that the operator describing the joint state has to be positive, but only that it has to be positive with respect to local measurements. A consistent theory that is not ruled out would allow the state to be negative with respect to some joint measurements (the Bell basis measurement, for example), but would not allow such measurements. Furthermore, the assumption that local operations commute and the global state assumption are both implicit in Fuchs' analysis. Without the latter, the possibility remains that there are extra degrees of freedom, not accessible via local measurements, that are not described by an operator on the tensor product Hilbert space.

It follows that Fuchs' conclusion is not stronger than the tensor product rule derived in this paper. The latter may be regarded as a generalization of Fuchs' proof to the case in which the subsystems A and B are not necessarily quantum.

C. Information theory, GNST and GLT

In addition to describing general properties of the framework, I investigated in detail two particular theories, GNST and GLT. I focussed on the information processing possibilities in these theories. One of the most interesting things to have emerged is that there is a tradeoff between the states of a theory and the allowed dynamics. This arises for the simple reason that an allowed transformation must take allowed states into allowed states. Thus the dynamics of both GNST and GLT is very simple for single systems. In GNST, a similar result holds for the simplest kind of bipartite system. The surprising consequence is that GNST is less powerful than quantum theory in many ways, despite including superquantum correlations. For example, teleportation and superdense coding are impossible. It is already clear that computation in GLT can be simulated efficiently classically, while the computational power of GNST remains open. Another open question is whether secure bit commitment is possible in either theory. Despite these remarks, it is surprising how many features of quantum theory have analogs in GNST. These obviously include the generic features demonstrated in Sec. V, along with entanglement and nonlocality. But they also include things I have not discussed in detail, such as the distinction between sharp and unsharp measurements, and preparation contextuality. (Other authors have also found features of quantum theory reproduced in other contexts. Masanes *et al.* [13] show that various features, including a no-cloning theorem, are present in all theories that are nonlocal and nonsignaling. Spekkens has introduced a toy theory that contains a remarkably wide range of quantum phenomena [11], although note that this theory is not contained in our framework as it does not allow arbitrary convex combinations of states.)

As mentioned above, one of the motivations of this work is to stimulate the study of computation in models that are more general than quantum theory. Some authors have already considered computation in nonstandard theories. However, these theories are often modifications of quantum theory that appear to have both unphysical consequences and immense computational power. It is suspected that quantum theory with a nonlinear Schrödinger equation is very powerful, enabling the solution of NP-complete problems in polynomial time, for example.[14] Aaronson has considered various modifications of quantum theory, including a model that assumes the ability to postselect measurement outcomes, and a hidden variable model in which the history of hidden states can be read out by the observer [5,6]. Various authors have considered classical and quantum computation in the presence of closed timelike curves [7,8]. Most recently, Aaronson and Watrous [9] have shown that BQP (Bounded Quantum Polynomial, the class of problems efficiently solvable by a quantum computer), in the presence of closed timelike curves, is equivalent to PSPACE (Polynomial Space, the class of problems solvable by a classical computer with polynomial memory). The framework introduced in this paper is the natural place to investigate computation in theories that are different from quantum theory, yet not obviously physically unreasonable or immensely powerful. I suggest that NP-complete problems cannot be solved efficiently by any theory in the framework. I also raise the following.

Conjecture 2. A quantum computer can simulate computation in any other theory in the framework with at most polynomial overhead.

The intuition behind this is that quantum theory achieves in some sense an optimal balance of allowed states and dynamics.

D. Interpretation

On the face of it, many theories that can be written down in the present framework have similar interpretive issues as

[14] In Ref. [3], it is claimed that nonlinear quantum theory can solve NP-complete and even # P-complete problems efficiently. (See Ref. [13] and references therein for a definition of the complexity class # P.) Aaronson complains [4] that in this particular case it is difficult to evaluate whether exponential precision is required.

quantum theory, if one tries to understand them in a way that goes beyond the purely operational. Consider a universe in which some theory other than quantum or classical (GNST perhaps) is verified in laboratory experiments. The denizens of such a universe would be having debates in many ways similar to the debates that surround quantum theory. Is a pure state better understood as a complete description of individual reality, as representing an ensemble, or as representing the degrees of belief of some agent?

Suppose that the inhabitants of this universe attempt to extend the theory to include a description of the measuring apparatus, and of the interaction between system and apparatus. This is always possible in classical and quantum theory. In quantum theory, this fact is expressed in the idea that the *Heisenberg cut* can be moved upwards indefinitely. Are classical and quantum theories special in this regard, or can this be done in any theory?

Even when the inhabitants succeed in constructing a measurement theory along these lines, it is plausible that many theories will have a *measurement problem*. In these theories, the system and apparatus are typically in some entangled state after interaction. Some inhabitants may suggest hidden variables or some kind of collapse dynamics. Does any theory admit an Everettian interpretation, or is there a special feature of quantum theory that is necessary for this to work?

I will not discuss these issues any further. I have raised them hoping that considering interpretive issues in a framework more general than quantum theory might give a new lease of life to the quantum debates.

E. Axioms

Aside from Hardy's derivation [14], what different ways are there of uniquely identifying quantum theory from the other theories in the framework by adding as few extra assumptions as possible? Several have pushed the idea that a quantum state is best understood as a summary of an agent's degrees of belief about the outcomes of future measurements on a system [32,57,58]. From this standpoint, Fuchs has argued that the formalism of quantum theory should be understood as a constraint on these degrees of belief, hopefully to be derived via a small number of postulates, along with an argument that any rational agent must accept [15]. Spekkens has also argued for an epistemic constraint as a foundational principle for quantum theory, although for Spekkens, beliefs are about underlying ontic states of a system rather than future measurement outcomes [11].

Clifton, Bub, and Halvorson (CBH) have taken a different approach and derived at least part of quantum theory from the assumption of (i) a no-signaling principle, (ii) a no-broadcasting principle, and (iii) the impossibility of secure bit commitment [59].[15] CBH assume a C^*-algebraic framework, which is broad enough to include classical theories and quantum theory, but is not as broad as the framework presented here. An open question is whether something like CBH's proof would go through in the broader framework, or whether there is some theory (GNST perhaps) that satisfies (i)–(iii) and is clearly not quantum.

Note added. Related independent work has appeared recently in Ref. [61].

ACKNOWLEDGMENTS

Research at Perimeter Institute is supported in part by the Government of Canada through NSERC and by the Province of Ontario through MEDT. I should very much like to thank Howard Barnum, Lucien Hardy, Nick Jones, Matthew Leifer, Lluis Masanes, Stefano Pironio, Sandu Popescu, Valerio Scarani, Tony Short, Rob Spekkens, and Alex Wilce for useful discussions.

APPENDIX A: PROOF OF LINEARITY OF TRANSFORMATIONS

The proof in this appendix is adapted from that of Hardy in Ref. [14]. It is included to keep this work self-contained.

A transformation is a map from allowed states of a system to allowed states. The map satisfies Eq. (7), reproduced here:

$$f\left(\sum_i q_i\vec{P}_i\right) = \sum_i q_i f(\vec{P}_i)\ \forall\ \ P_i \in \mathcal{S},$$

$$\text{for } 0 \leqslant q_i \leqslant 1, \quad \sum_i q_i = 1. \tag{A1}$$

The map should also satisfy

$$f(\vec{0}) = \vec{0}.$$

[This follows from the interpretation of unnormalized states. Recall that if a particular outcome i of some operation occurs with probability $q<1$, then we associate with that outcome an unnormalized vector $\vec{P}$. Each entry of $\vec{P}$ gives the joint probability of obtaining outcome i for the original operation, and outcome j for a fiducial measurement performed immediately afterwards. Thus if $q=0$, it follows that the associated $\vec{P}=\vec{0}$. By definition, an entry in the vector $f(\vec{0})$ represents the joint probability of getting the following outcomes in sequence: outcome i for the original operation, then whatever outcome it is that corresponds to the transformation f, and then outcome j for a fiducial measurement. But these probabilities must all be zero if the probability of outcome i is zero.]

Writing the first of the above equations with $i=1,2$, and setting $\vec{P}_2=\vec{0}$, gives

$$f(q\vec{P}) = qf(\vec{P}) \quad \forall\ \vec{P} \in \mathcal{S}, \quad \text{for } 0 \leqslant q \leqslant 1.$$

Suppose that $\vec{P}$ is a pure state $\in\mathcal{S}$. Pure states are by definition normalized. If $r>1$, then $f(r\vec{P})$ is initially undefined because $r\vec{P} \notin \mathcal{S}$, so we are free to stipulate that

$$f(r\vec{P}) = rf(\vec{P}) \quad \forall\ \vec{P} \in \mathcal{S}, \quad r \geqslant 0.$$

Define $\mathcal{S}_+$ as the set of all vectors that can be written in the form $r\vec{P}$ with $\vec{P}\in\mathcal{S}$ and $r\geqslant 0$. It is a convex cone [60]. Equation (A1) can be extended slightly:

[15] Whether they succeed in deriving the full structure of quantum theory is debatable. But they do establish the existence of noncommuting measurements and of entanglement.

$$f\left(\sum_i r_i\vec{P}_i\right)=\sum_i r_i f(\vec{P}_i) \qquad \forall\, P_i\in\mathcal{S}_+ \quad \text{for } r_i\geqslant 0. \qquad \text{(A2)}$$

Now suppose that

$$\vec{P}=\sum_i s_i\vec{P}_i, \qquad \text{(A3)}$$

where $\vec{P},\vec{P}_i\in\mathcal{S}_+$, and the s_i are real. Let $i\in A_-$ if $s_i<0$ and $i\in A_+$ if $s_i\geqslant 0$. Rewrite Eq. (A3) as

$$\vec{P}+\sum_{i\in A_-}|s_i|\vec{P}_i=\sum_{i\in A_+}s_i\vec{P}_i.$$

Each side is a conic combination of vectors in $\mathcal{S}_+$, thus Eq. (A2) applies, and rearranging we get

$$f(\vec{P})=\sum_i s_i f(\vec{P}_i).$$

Finally, for any vector $\vec{Q}\notin\mathcal{S}_+$, $f(\vec{Q})$ can be defined uniquely by linear extension if $\vec{Q}$ lies in the subspace spanned by $\mathcal{S}$. The action of f on the rest of the vector space is arbitrary but may be defined to be linear. ■

APPENDIX B: DERIVATION OF TENSOR PRODUCT RULE

As discussed in the main text, the state of a joint system AB can be written

$$\vec{P}^{AB}\equiv\left(\begin{array}{c} P(a=1,b=1|X=1,Y=1)\\ P(a=1,b=2|X=1,Y=1)\\ \vdots\\ \hline P(a=1,b=1|X=1,Y=2)\\ P(a=1,b=2|X=1,Y=2)\\ \vdots\\ \hline \vdots \end{array}\right).$$

Proof of Theorem 1. This theorem is trivial. Let $\vec{P}^{AB}\in V^{AB}$, $\vec{P}^{A}\in V^{A}$, and $\vec{P}^{B}\in V^{B}$. Define the vector $\vec{Q}^{AB}_{ijkl}$ as the vector with a 1 for the entry corresponding to the joint outcome ij of the joint fiducial measurement kl, and 0s elsewhere. Similarly $\vec{Q}^{A}_{ik}$ and $\vec{Q}^{B}_{jl}$. Now identify $\vec{Q}^{AB}_{ijkl}$ with $\vec{Q}^{A}_{ik}\otimes\vec{Q}^{B}_{jl}$ and extend linearly. ■

Proof of Theorem 2. Consider a joint system AB. For each of the fiducial measurements that define the state of system B, there must be at least one operation on the joint system AB that corresponds to performing that measurement. Let this operation for the jth fiducial measurement be characterized by the set of matrices $\{M_{ij}\}$, where there is a value of i for each outcome and j is fixed. When the transformation M_{ij} acts on AB, the resulting state is the unnormalized state $\vec{P}^{AB}_{ij}\in\mathcal{S}^{AB}$. The corresponding reduced state for A is the unnormalized state $\vec{P}^{A}_{ij}$. By Constraint 2, $\vec{P}^{A}_{ij}\in\mathcal{S}^{A}$. If a fiducial measurement is now performed on A, the state $\vec{P}^{A}_{ij}$ gives the (unnormalized) probabilities for the different outcomes. It follows that $\vec{P}^{AB}$ can be written in the form

$$\vec{P}^{AB}=\sum_{ij}\vec{P}^{A}_{ij}\otimes\vec{Q}^{B}_{ij}, \qquad \text{(B1)}$$

with $\vec{P}^{A}_{ij}\in\mathcal{S}^{A}$ and $\vec{Q}^{B}_{ij}$ as above. Now consider a vector $\vec{U}\otimes\vec{W}\in V^{AB}$, with $\vec{W}\in\mathcal{S}^{B}$ but $\vec{U}\perp\mathcal{S}^{A}$, where this means that $\vec{U}$ is orthogonal to all vectors in $\mathcal{S}^{A}$. From Eq. (B1) it follows that $(\vec{U}\otimes\vec{W})\cdot\vec{P}^{AB}=0$. A similar result holds if $\vec{W}\perp\mathcal{S}^{B}$ and $\vec{U}\in\mathcal{S}^{A}$. Thus $\vec{P}^{AB}$ lies in the subspace of V^{AB} that is spanned by vectors from $\mathcal{S}^{A}\otimes\mathcal{S}^{B}$. Equation (19) follows. The vectors on the right-hand side of this equation can be assumed normalized, since any multiplying factor can be subsumed into the corresponding r_i. They can be assumed pure, since a mixed state can always be expressed as a convex combination of pure states and $\vec{0}$. But any term with $\vec{0}$ will not contribute. Theorem 2 follows. ■

Proof of Theorem 3. Consider a joint system AB and a transformation T^{A} of system A alone. T^{A} corresponds to a matrix M^{A} such that $\vec{P}^{A}\to\vec{P}'^{A}=M^{A}\cdot\vec{P}^{A}$. The aim is to determine the effect of this transformation on the joint state $\vec{P}^{AB}$. From Sec. II A, this will correspond to a matrix $\tilde{M}^{A}$ such that $\vec{P}^{AB}\to\vec{P}'^{AB}=\tilde{M}^{A}\cdot\vec{P}^{AB}$. But what is the relation between M^{A} and $\tilde{M}^{A}$?

Consider the following procedure. First, the transformation T^{A} is applied. Then fiducial measurements are performed on systems A and B. The (unnormalized) joint probabilities for the outcomes of these measurements are then the entries of the vector $\vec{P}'^{AB}$. However, by Assumption 4, the ordering of operations on systems A and B does not matter. Thus the following procedure is equivalent. First, a fiducial measurement is performed on system B. Note that the reduced state of system A conditioned on a particular outcome for this measurement is defined by the vector $\vec{P}^{AB}$. Next, the transformation T^{A} is performed on system A. Finally, a fiducial measurement is performed on system A.

In the second procedure, we know how to apply the transformation T^{A}, since it is enough to consider its action on system A alone, and we know that $\vec{P}^{A}\to\vec{P}'^{A}=M^{A}\cdot\vec{P}^{A}$. We obtain

$$\vec{P}'^{AB}_{ijkl}=\sum_{i'k'}(M^{A})_{ik;i'k'}\vec{P}^{AB}_{i'jk'l}=\sum_{i'k'j'l'}(M^{A})_{ik;i'k'}\delta_{jj'}\delta_{ll'}\vec{P}^{AB}_{i'j'k'l'}.$$

But

$$(M^{A}\otimes I^{B})_{ijkl;i'j'k'l'}=(M^{A})_{ik;i'k'}\delta_{jj'}\delta_{ll'}$$

thus

$$\vec{P}'^{AB}=(M^{A}\otimes I)\cdot\vec{P}^{AB}.$$

This holds for all $\vec{P}^{AB}\in\mathcal{S}^{AB}$, and the action of T^{A} on vectors $\vec{P}^{AB}\notin\mathcal{S}^{AB}$ is arbitrary. It follows that we lose no generality in identifying

$$\tilde{M}^{A}=M^{A}\otimes I^{B}.$$

■

APPENDIX C: GENERIC FEATURES

This appendix contains proofs of the results of Sec. V.

Proof of Theorem 4. Consider a particular type of system in some theory. Suppose that the subspace spanned by allowed states of the system has dimension d and that every mixed state has a unique decomposition into pure states and $\vec{0}$. The only convex set with this property is a simplex with $d+1$ vertices. One of these vertices is the state $\vec{0}$. It is always possible to find an invertible linear transformation N such that the other vertices are transformed into the vectors (1,0,0,…,0), (0,1,0,…,0), and so on. Recall from Sec. III that if this transformation acts on the set $\mathcal{S}$, then the theory is not changed, since $\vec{R}^T \rightarrow \vec{R}^T \cdot N^{-1}$ and $M \rightarrow N \cdot M \cdot N^{-1}$ for measurements and transformations. If the state space for every type of system, including multipartite systems, is of this form, then the theory is classical. ■

Proof of Theorem 5. Consider a system with a set of allowed states $\mathcal{S}$, spanning a subspace $V_{\mathcal{S}}$, and let d be the dimension of $V_{\mathcal{S}}$. Choose a set of d distinct pure states $\{\vec{P}_1, \ldots, \vec{P}_d\}$ that are linearly independent and collectively span $V_{\mathcal{S}}$. Suppose that a particular transformation is nondisturbing. Its action on each of the $\vec{P}_i$ is given by $M \cdot \vec{P}_i = c_i \vec{P}_i$ with $0 \leqslant c_i \leqslant 1$. If $\mathcal{S}$ is a simplex, then the set $\{\vec{P}_1, \ldots, \vec{P}_d\}$ must contain all the pure states. Since the $\vec{P}_i$ are linearly independent, the c_i can be chosen independently without contradiction. For any other type of system, there are at least $d+1$ pure states. Consider a pure state $\vec{Q}$ that is not contained in the set $\{\vec{P}_1, \ldots, \vec{P}_d\}$. If the transformation is nondisturbing, then $M \cdot \vec{Q} = e\vec{Q}$ with $0 \leqslant e \leqslant 1$. Since $\{\vec{P}_1, \ldots, \vec{P}_d\}$ is a basis for $V_{\mathcal{S}}$, $\vec{Q}$ has a unique decomposition of the form $\vec{Q} = \Sigma_i d_i \vec{P}_i$, where at least two of the d_i are nonzero. If d_j and d_k are nonzero, then $c_j = c_k = e$. Thus M acts as e times the identity on the subspace of $V_{\mathcal{S}}$ spanned by $\vec{P}_j$ and $\vec{P}_k$. By repeating this reasoning for every pure state $\vec{Q}$, the set $\{\vec{P}_1, \ldots, \vec{P}_d\}$ can be divided into subsets such that (i) if $\vec{P}_j$ and $\vec{P}_k$ are in the same subset, then $c_j = c_k$ for any nondisturbing transformation, and (ii) if $\vec{P}_j$ and $\vec{P}_k$ are in different subsets then there is no pure state $\vec{Q}$ such that both d_j and d_k are nonzero. Each subset defines a subspace V_i of $V_{\mathcal{S}}$ and the theorem follows. ■

Proof of Theorem 6. Theorem 6 is proven using Theorem 5. We show that if there is a probabilistic universal cloning procedure, then for any two pure states $\vec{P}_1$ and $\vec{P}_2$, there is a nondisturbing transformation M' such that $|M' \cdot \vec{P}_1| \neq |M' \cdot \vec{P}_2|$. This in turn implies that $\mathcal{S}$ is a simplex. If $\mathcal{S}$ is a simplex for every type of system, then the theory is classical.

Suppose that there is a standard state $\vec{Q}$ and a transformation M such that for each pure state $\vec{P}$, $M \cdot (\vec{P} \otimes \vec{Q}) = c\vec{P} \otimes \vec{P}$. The number c may vary with $\vec{P}$ but is >0 for all $\vec{P}$. Consider a procedure in which a system is in the state $\vec{P}_1$ or $\vec{P}_2$, an ancilla is added in the standard state $\vec{Q}$, and the cloning operation $\{M, F\}$ performed on the joint system. The transformation M corresponds to the success outcome and F to the fail outcome. If $\vec{P}_1$ and $\vec{P}_2$ are different states there must be some operation $\{N_1, N_2\}$ such that $|N_1 \cdot \vec{P}_1| \neq |N_1 \cdot \vec{P}_2|$. If cloning succeeded, perform this operation on the ancilla. Output the result and throw away the ancilla.

This entire procedure may be regarded as an operation on the system alone (see the remarks following Assumption 7). It can be written $O' = \{M_1', M_2', F'\}$, where M_1' corresponds to successful cloning followed by the N_1 outcome, M_2' corresponds to successful cloning followed by the N_2 outcome, and F' corresponds to failed cloning. By construction, each of M_1' and M_2' is nondisturbing and $|M_i' \cdot \vec{P}_1| \neq |M_i' \cdot \vec{P}_2|$ for at least one of $i = 1, 2$. ■

APPENDIX D: DYNAMICS IN GNST AND GLT

This appendix contains proofs of Theorems 7, 8, and 9, all of which concern dynamics in GNST or GLT.

Proof of Theorem 7. This theorem concerns transformations of single systems in either GNST or GLT. A transformation of an (n,k) system can be written

$$\begin{pmatrix} P'(a=1|X=1) \\ \vdots \\ P'(a=k|X=1) \\ \hline \vdots \\ \hline P'(a=1|X=n) \\ \vdots \\ P'(a=k|X=n) \end{pmatrix} = \left(\begin{array}{c|c|c} M_{11} & \cdots & M_{1n} \\ \hline \vdots & & \vdots \\ \hline M_{n1} & \cdots & M_{nn} \end{array}\right) \times \begin{pmatrix} P(a=1|X=1) \\ \vdots \\ P(a=k|X=1) \\ \hline \vdots \\ \hline P(a=1|X=n) \\ \vdots \\ P(a=k|X=n) \end{pmatrix}. \tag{D1}$$

The transformation matrix is M, an $nk \times nk$ matrix. If the fiducial measurement $X=1$ has k outcomes, then the top k rows of this matrix determine the probabilities of outcomes when the $X=1$ measurement is performed on the transformed state $\vec{P}'$. Denote the $k \times nk$ submatrix consisting of these rows M_1. The next k rows are associated with the fiducial measurement $X=2$, so denote the corresponding submatrix by M_2, and so on. The first k columns of M_i multiply into those components of $\vec{P}$ that correspond to the fiducial measurement $X=1$ being performed. Denote the $k \times k$ subsubmatrix consisting of these columns M_{i1}. Similarly M_{i2}, and so on. Note that each row in M, considered as a vector $\vec{R}$, must represent a possible measurement outcome. This is because if the transformation acts on a state $\vec{P}$, then $\vec{R} \cdot \vec{P}$ gives the corresponding entry in the transformed state $\vec{P}'$, which must be between 0 and 1 for all $\vec{P} \in \mathcal{S}$. Furthermore, when the transformation is normalization-preserving, the rows $\vec{R}_j$ from a particular M_i satisfy $\Sigma_j \vec{R}_j \cdot \vec{P}=1$, whenever $\vec{P}$ is normalized. Hence the rows from a particular M_i correspond to a multiple-outcome measurement. One way of performing this measurement is simply to perform the transformation M first, and then to perform fiducial measurement $X=i$.

There is some redundancy in a measurement vector $\vec{R}$, and in the matrix M. If $\vec{R}\cdot\vec{P}=\vec{R}'\cdot\vec{P}\,\forall\ \vec{P}\in\mathcal{S}$, then $\vec{R}$ and $\vec{R}'$ represent the same measurement outcome. In particular, if $\vec{R}'=\vec{R}+\vec{C}$, where $\vec{C}\cdot\vec{P}=0\,\forall\ \vec{P}\in\mathcal{S}$, then $\vec{R}$ and $\vec{R}'$ represent the same measurement. An example of such a $\vec{C}$ is

$$\vec{C}=(1,\ldots,1|-1,\ldots,-1|0,\ldots,0|\ldots),$$

where $\vec{C}\cdot\vec{P}=0\,\forall\ \vec{P}\in\mathcal{S}$ is ensured by the normalization of $\vec{P}$. The first step in the proof is to show that any $\vec{R}$ is equivalent in this sense to an $\vec{R}'$ with all components $\geqslant 0$.

For this, consider the set of allowed normalized states. This is precisely the set of vectors satisfying the conditions

$$\sum_i P(a=i|X=j)=\sum_i P(a=i|X=k)\ \ \forall\ \ j,k, \tag{D2}$$

$$P(a=i|X=j)\geqslant 0\ \ \forall\ \ i,j, \tag{D3}$$

$$\sum_i P(a=i|X=1)=1. \tag{D4}$$

Define $\mathcal{S}_+$ as the set of vectors of the form $r\vec{P}$, with $r\geqslant 0$ and $\vec{P}\in\mathcal{S}$, and note that in the case of GNST or GLT, $\mathcal{S}_+$ is a polyhedral cone [60]. It can also be defined as the set of vectors satisfying conditions (D2) and (D3). The defining inequalities (D3) can each be written in the form $\vec{C}_i\cdot\vec{P}\geqslant 0$, where $\vec{C}_i$ is a constant vector with a 1 in the ith position and 0s elsewhere. The equalities (D2) can each be written as the conjunction of two inequalities: $\vec{D}_j\cdot\vec{P}\geqslant 0$ and $\vec{D}_j\cdot\vec{P}\leqslant 0$ for some constant D_j. Define $\mathcal{R}_+$ as the set of vectors $\vec{R}$ such that $\vec{R}\cdot\vec{P}\geqslant 0\,\forall\ \vec{P}\in\mathcal{S}_+$. This is the set of unnormalized measurements and is the dual cone to $\mathcal{S}_+$. It can be shown that if a polyhedral cone is defined by $\{\vec{P}:\vec{A}_i\cdot\vec{P}\geqslant 0\,\forall\ i\}$, then the dual cone is equal to the conic hull of the vectors $\vec{A}_i$. Thus elements of $\mathcal{R}_+$ can be written

$$\vec{R}=\sum_i \lambda_i\vec{C}_i+\sum_j \mu_j\vec{D}_j, \tag{D5}$$

where $\lambda_i\geqslant 0$ and μ_j can be positive or negative. Finally, the vectors $\vec{D}_j$ all satisfy $\vec{D}_j\cdot\vec{P}=0\,\forall\ \vec{P}\in\mathcal{S}_+$. Hence any $\vec{R}$ of this form is equivalent to an $\vec{R}$ of the form

$$\vec{R}=\sum_i \lambda_i\vec{C}_i, \tag{D6}$$

and without loss of generality, the components of $\vec{R}$ can be assumed $\geqslant 0$. This applies both to $\vec{R}$ considered as a measurement outcome and to $\vec{R}$ considered as a row of a transformation matrix M.

Assume, then, that M is written in a form with all entries $\geqslant 0$. To conclude the proof, note that M acting on any properly normalized state [satisfying both Eqs. (D2) and (D4)] must result in a state that is also properly normalized. This implies the following. Consider the matrix M_{ij}. Denote the sum of the elements in the first column by S_1^{ij}, the sum of the elements in the second column by S_2^{ij}, and so on. Then $S_1^{ij}=S_2^{ij}=\cdots=S_k^{ij}$ and $\Sigma_j S_1^{ij}=1$. Hence the matrix M_{ij} is of the form α_{ij} times a stochastic matrix, with $0\leqslant\alpha_{ij}\leqslant 1$ and $\Sigma_j\alpha_{ij}=1$. One may easily check that any transformation that is equivalent to a procedure of the form of Fig. 5 is represented by a matrix of this form with $\alpha_{ik}=1$ for some k and $\alpha_{ij}=0$ for $j\neq k$. Hence we have obtained the general result that any allowed M is a convex combination of transformations of the form of Fig. 5. ■

Proof of Theorem 8. Let an m-outcome measurement on an (n,k) system have outcomes corresponding to $\vec{R}_1,\ldots,\vec{R}_m$, and construct the $m\times nk$ matrix

$$N=\begin{pmatrix}\vec{R}_1^T\\ \vdots\\ \vec{R}_m^T\end{pmatrix}.$$

Denote the submatrix consisting of the first k columns of N by N_1, that consisting of the next k columns by N_2, and so on. The same arguments as in the proof of Theorem 7 can be used to establish that N can be chosen such that all entries are $\geqslant 0$. Then use the fact that $\Sigma_i\vec{R}_i\cdot\vec{P}=1$ for normalized $\vec{P}$, and arguments similar to those in the proof of Theorem 7, to establish that $N_i=\alpha_i S_i$ for $0\leqslant\alpha_i\leqslant 1$, $\Sigma_i\alpha_i=1$, and S_i stochastic. The theorem follows. ■

Proof of Theorem 9. Begin as before by showing that without loss of generality, the matrix M can be taken to have all entries $\geqslant 0$. This part of the proof is identical, except that to conditions (D2)–(D4), one should add the no-signaling constraints

$$\sum_j P(a=i,b=j|X=k,Y=1)=\sum_j P(a=i,b=j|X=k,Y=2)\ \ \forall\ \ i,k \tag{D7}$$

$$\sum_i P(a=i,b=j|X=1,Y=l)=\sum_i P(a=i,b=j|X=2,Y=l)\ \ \forall\ \ j,l. \tag{D8}$$

Like the conditions (D2), these constraints can be written as the conjunction $\vec{D}_j\cdot\vec{P}\geqslant 0$ and $\vec{D}_j\cdot\vec{P}\leqslant 0$, and $\vec{R}$ can be written in the form of Eq. (D5), hence in the form of Eq. (D6). Now impose that $\vec{P}'=M\cdot\vec{P}$ is normalized for any allowed normalized $\vec{P}$, that is any $\vec{P}$ that satisfies conditions (D2)–(D4), (D7), and (D8). Proving that any such M represents a convex combination of transformations of the form of Fig. 7 (or the reversed form with respect to the two subsystems) is a tedious brute force exercise that is omitted. As with Theorem 8, the proof of Theorem 10 is a straightforward variation. ■

[1] Workshop titled "What is Quantum in Quantum Computing?", Konstanz, Germany, May 19–20, 2005.

[2] M. Troyer and U.-J. Wiese, Phys. Rev. Lett. **94**, 170201 (2005).

[3] D. S. Abrams and S. Lloyd, Phys. Rev. Lett. **81**, 3992 (1998).

[4] S. Aaronson, SIGACT News **36**(1), 30 (2005).

[5] S. Aaronson, Phys. Rev. A **71**, 032325 (2005).

[6] S. Aaronson, Proc. R. Soc. London, Ser. A **461**, 3473 (2005).

[7] T. A. Brun, Found. Phys. Lett. **16**, 245 (2003).

[8] D. Bacon, Phys. Rev. A **70**, 032309 (2004).

[9] S. Aaronson and J. Watrous (unpublished).

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

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

[12] J. A. Smolin, Quantum Inf. Comput. **5**, 161 (2005).

[13] Ll. Masanes, A. Acin, and N. Gisin, Phys. Rev. A **73**, 012112 (2006).

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

[15] C. A. Fuchs, in *Quantum Theory: Reconstruction of Foundations*, edited by A. Yu. Khrennikov (Växjö University Press, Växjö, 2002), p. 463.

[16] L. A. Khalfi and B. S. Tsirelson, in *Symposium on the Foundations of Modern Physics*, edited by P. Lahti and P. Mittelstaedt (World Scientific, Singapore, 1985), pp. 441–460.

[17] S. Popescu and D. Rohrlich, Found. Phys. **24**, 379 (1994).

[18] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts, Phys. Rev. A **71**, 022101 (2005).

[19] S. Wolf and J. Wullschleger, in *Proceedings of International Symposium on Information Theory* (ISIT) 2005.

[20] A. J. Short, N. Gisin, and S. Popescu, Quantum Inf. Process. **5**, 131 (2006).

[21] H. Buhrman, M. Christandl, F. Unger, S. Wehner, and A. Winter, Proc. R. Soc. London, Ser. A **462**, 1919 (2006).

[22] A. Broadbent and A. A. Méthot, Theor. Comput. Sci. **358**, 3 (2006).

[23] J. Barrett and S. Pironio, Phys. Rev. Lett. **95**, 140401 (2005).

[24] N. S. Jones and Ll. Masanes, Phys. Rev. A **72**, 052312 (2005).

[25] W. van Dam, e-print quant-ph/0501159.

[26] G. Brassard, H. Buhrman, N. Linden, A. A. Méthot, A. Tapp, and F. Unger, Phys. Rev. Lett. **96**, 250401 (2006).

[27] W. K. Wootters, Found. Phys. **16**, 391 (1986).

[28] P. G. L. Mana, e-print quant-ph/0305117.

[29] P. G. L. Mana, in *Quantum Theory: Reconsideration of Foundations 2*, edited by A. Yu. Khrennikov (Växjö University Press, Växjö, 2004), p. 387.

[30] E. C. G. Stueckelberg, Helv. Phys. Acta **33**, 727 (1960).

[31] W. K. Wootters, in *Complexity, Entropy and the Physics of Information*, edited by W. H. Zurek (Addison-Wesley, Reading, MA, 1990).

[32] C. M. Caves, C. A. Fuchs, and R. Schack, J. Math. Phys. **43**, 4537 (2002).

[33] R. F. Werner, Phys. Rev. A **40**, 4277 (1989).

[34] J. Barrett, Phys. Rev. A **65**, 042302 (2002).

[35] F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).

[36] B. S. Cirel'son, Lett. Math. Phys. **4**, 93 (1980).

[37] H. Barnum, C. A. Fuchs, J. M. Renes, and A. Wilce, e-print quant-ph/0507108.

[38] H. Barnum, J. Barrett, M. Leifer, and A. Wilce (unpublished).

[39] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, e-print quant-ph/0611295.

[40] J. Barrett and M. Leifer (unpublished).

[41] J. Barrett (unpublished).

[42] R. W. Spekkens, Phys. Rev. A **71**, 052108 (2005).

[43] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, Phys. Rev. Lett. **70**, 1895 (1993).

[44] C. H. Bennett and S. J. Wiesner, Phys. Rev. Lett. **69**, 2881 (1992).

[45] A. K. Ekert, Phys. Rev. Lett. **67**, 661 (1991).

[46] C. H. Bennett and G. Brassard, in *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing* (IEEE, New York, 1984), p. 175.

[47] S. Even, O. Goldreich and A. Lempel, in *Proceedings CRYPTO 82*, edited by R. L. Rivest, A. Sherman, and D. Chaum (Plenum, New York, 1982), p. 205.

[48] H.-K. Lo, Phys. Rev. A **56**, 1154 (1997).

[49] G. Brassard, C. Crépeau, and J.-M. Robert, in *27th Symposium of Foundations of Computer Science* (IEEE, New York, 1986), p. 168.

[50] J. Killian, in *Proceedings of 20th ACM Symposium on Theory of Computing* (ACM, New York, 1988), p. 20.

[51] R. Cleve, W. van Dam, M. Nielsen, and A. Tapp, in *Proceedings of the first NASA International Conference on Quantum Computing and Quantum Communication*, LNCS Vol. 1509 (Springer-Verlag, Heidelberg, 1998), p. 61.

[52] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani, e-print quant-ph/9804043.

[53] G. Svetlichny, e-print quant-ph/0410036.

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

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

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

[57] C. M. Caves, C. A. Fuchs, and R. Schack, Phys. Rev. A **65**, 022305 (2002).

[58] C. M. Caves, C. A. Fuchs, and R. Schack, e-print quant-ph/0608190.

[59] R. Clifton, J. Bub, and H. Halvorson, Found. Phys. **33**, 1561 (2003).

[60] A. Barvinok, *A Course in Convexity*, Graduate Studies in Mathematics Vol. 54 (American Mathematical Society, Providence, RI, 2002).

[61] A. J. Short, S. Popescu, and N. Gisin, Phys. Rev. A **73**, 012101 (2006).