Gleason-Busch theorem for sequential measurements

Kieran Flatt,* Stephen M. Barnett, and Sarah Croke
School of Physics and Astronomy, University of Glasgow, Glasgow, G12 8QQ, UK

Gleason's theorem is a statement that, given some reasonable assumptions, the Born rule used to calculate probabilities in quantum mechanics is essentially unique. We show that Gleason's theorem contains within it also the structure of sequential measurements, and along with this the state update rule. We give a small set of axioms, which are physically motivated and analogous to those in Busch's proof of Gleason's theorem, from which the familiar Kraus operator form follows. An axiomatic approach has practical relevance as well as fundamental interest, in making clear those assumptions which underlie the security of quantum communication protocols. The two time formalism is seen to arise naturally.

I. INTRODUCTION

The Born rule is fundamental to quantum mechanics, giving a prescription for alternatively predicting or interpreting measurement statistics. It is perhaps natural to wonder, therefore, whether the structure of quantum mechanics allows any other rule for calculating probabilities, and it can be shown that it does not. Gleason [1] showed that, within quantum theory (that is, assuming that measurements are described by projectors, and given some reasonable assumptions that any probability measure must obey) every probability allowed by quantum mechanics is calculated by a trace rule:

$$\mathrm{P}(i) = \mathrm{Tr}(\hat{\rho}\hat{P}_i). \tag{1}$$

Busch [2] generalized Gleason's theorem in two important ways: Gleason's original proof applied only to systems of dimension 3 or more, and also assumed that measurements were described by projectors. Busch's proof assumed only that measurements were described by positive operators, thus including the more general POVM (positive operator-valued measure, also known as POM or probability operator measure) formalism [3]. This was later generalised further: Busch's proof applies to complete measurements, for which the operators sum to the identity, a restriction which was relaxed in [4]. This generalisation means that probability rules may be derived rather directly for cases involving post-selection, and for retrodiction, for example.

In this work we are concerned with sequential measurements: in quantum mechanics measurement causes disturbance, and the state of the system must be updated post-measurement. This state update rule is given for projective measurements by the von Neumann projection postulate [5] or Lüders rule [6]:

$$\hat{\rho} \to \frac{\hat{P}_i\hat{\rho}\hat{P}_i}{\mathrm{Tr}(\hat{\rho}\hat{P}_i)}, \tag{2}$$

or more generally by the Kraus operator formalism [7]:

$$\hat{\rho} \to \frac{\sum_k \hat{A}_{ik}\hat{\rho}\hat{A}_{ik}^\dagger}{\mathrm{Tr}(\hat{\rho}\sum_k \hat{A}_{ik}^\dagger\hat{A}_{ik})}. \tag{3}$$

The description of measurement via positive operators is thus only part of the story, for a complete description of measurement we require both a means of calculating measurement statistics, and of expressing the change in state. The Busch-Gleason theorem, which takes as an assumption that measurements are described by positive operators, thus does not immediately lend itself to sequential measurement. We show nevertheless that a joint probability measure on pairs of measurements may be derived via an extension of Gleason's theorem, which recovers the usual Kraus form for sequential measurement.

The structure of transformations in quantum theory is, of course, well understood: all physically allowed transformations are described by so-called completely positive maps. These have several equivalent representations: the Kraus form [7], and the Choi-Jamiolkowski isomorphisms [8, 9], each of which may be derived from the usual structure of quantum mechanics on Hilbert space (see e.g. [10]). The advantage of an axiomatic approach is to make clear exactly on which assumptions this structure relies, an approach of both fundamental interest, and of practical relevance. In the era of quantum communication and security it is crucial to know which aspects of quantum theory are required for security of such schemes, both to feed into security proofs and to reassure users.

There is by now in the literature a long tradition of axiomatic approaches to both the description of measurement in quantum theory [2, 11–13], and indeed to derive the structure of quantum theory from simple principles [14–17]. We note in particular that previous work has addressed a similar scenario to that of interest here: Cassinelli and Zanghi [12] derived the Lüders rule for state update through consideration of conditional probabilities via Gleason type arguments. This is however not readily generalised to more general measurements, those which are not described by projectors. More recently Shrapnel et al [13], starting from an assumption that transformations are described by completely positive maps also used an axiomatic approach similar to Busch and Gleason to derive a probability measure which en-

* k.flatt.1@research.gla.ac.uk

compasses both the Born rule and state update rule. Motivated by recent work on indefinite causal order in quantum mechanics [18–21], Shrapnel et al's work derives the most general rule resulting in a probability measure on the set of completely positive maps. In the present work, by contrast, we show that sequential measurements correspond to completely positive maps and derive the most general form of these, from a few simple axioms.

Work on generalised probabilistic theories (we note in particular [14, 15]) takes as a starting point that probabilities may be expressed as an inner product of a vector describing the measurement and one describing the preparation. Transformations are described by operators on these vectors, and a key point, in common with the present work, is a proof that such transformations must further be *linear* operators on the space. A theory is then defined by the structure of the vector spaces describing states and measurements. Similarly, a common thread through the present work is that probabilities may be expressed as inner products, it is simply a case of defining the appropriate space in each case.

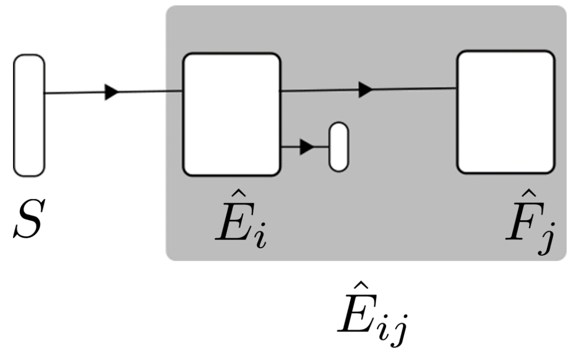

FIG. 1. A measurement procedure can be visualised using this flowchart. A preparation procedure, here labelled S, will output a quantum system, which according to the Gleason-Busch theorem can be described by a density matrix $\hat{\rho}$. The first measurement is associated with a set of effects $\hat{E}_i$, while the second measurement is associated with effects $\{\hat{F}_j\}$. Alternatively the whole procedure comprising both measurements is itself a measurement, and is represented by a set of effects $\{\hat{E}_{ij}\}$.

II. SEQUENTIAL MEASUREMENTS

We begin by summarising Busch's proof of Gleason's theorem, before discussing the extension to sequential measurements. Busch [2] assumes that measurement outcomes are associated with positive semi-definite operators $\hat{E}$ such that $\hat{E} \leq \hat{I}$ (referred to therein as effects). He then seeks the most general probability measure on this set of operators, $\nu(\hat{E})$, which satisfies the following propositions:

(P1) $0 \leq \nu(\hat{E}) \leq 1$.

(P2) $\nu(\hat{I}) = 1$.

(P3) $\nu(\hat{E} + \hat{F} + \ldots) = \nu(\hat{E}) + \nu(\hat{F}) + \ldots$

The proof proceeds by showing that the additivity proposition (P3) on effects may always be extended to linearity on all Hermitian operators: that is, for each such ν, we can define an extension which acts not only on positive operators, but on all Hermitian operators, and which is linear. The allowed ν are all real, according to proposition (P1), and noting that the set of Hermitian operators forms a real vector space, it therefore follows that each such function, by definition, is a vector in the dual space [22]. Thus every measure on effects may be associated with a Hermitian operator $\hat{\rho}$: $\nu(\hat{E}) = \mathrm{Tr}(\hat{\rho}\hat{E})$, using the Hilbert-Schmidt inner product $(\hat{A}, \hat{B}) = \mathrm{Tr}(\hat{A}^\dagger \hat{B})$. Propositions (P1) and (P2) further constrain $\hat{\rho}$ to be positive and trace one respectively.

We note at this point that the choice of Hilbert-Schmidt inner product is not unique, we can in principle choose any bi-linear form. The most general probability rule is thus given by $v(\hat{E}) = \mathrm{Tr}(\mathcal{L}(\hat{\rho})\hat{E})$ for some linear superoperator $\mathcal{L}$. Of course this doesn't give us any additional generality: the requirement now is that $\mathcal{L}(\hat{\rho})$ be a positive operator, and via the substitution $\mathcal{L}(\hat{\rho}) \to \hat{\rho}$ we recover the previous formulation. Indeed attempts to generalize quantum theory have resulted in theories with a non-standard choice of inner product [23, 24], later shown to be equivalent to standard quantum theory as long as the set of vectors which are allowed to represent states is updated accordingly [25–27]. We return however to the nonuniqueness of the inner product in the present context later.

In the sequential measurement case, we consider a setup like that shown in Figure 1, in which a single system undergoes two successive measurements. Following Busch, we take as a definition that measurements are represented by effects, positive operators $\hat{A}$ defined on a Hilbert space, and include entanglement by allowing for measurements to be performed on subspaces of those (i.e. $\hat{A} \otimes \hat{B}$ is an allowed effect). This is what is brought over from standard quantum mechanics; what is derived is the probabilistic structure. Although we note there is much recent interest in causally neutral formulations of quantum theory [19, 21] and non-fixed causal orderings [18, 20], for simplicity we consider here a fixed causal order: the measurements we consider are performed sequentially, and on the same system. Under the assumptions of the Gleason-Busch theorem, measurements are described by positive operators: we thus associate to the first measurement a set of positive operators $\{\hat{E}_i\}$, to the second measurement the set $\{\hat{F}_j\}$. The combination of measurements, of course, is itself a measurement procedure; we associate to this procedure the operators $\{\hat{E}_{ij}\}$. Our task is to derive a relationship between these three sets of operators.

We first note that, due to our choice of causal order, the statistics of the first measurement alone are independent of whether the second measurement is performed or not, and may be reconstructed by coarse-graining over the second measurement. Thus for any $\hat{E}_{ij}$ representing the joint measurement procedure we must have

$$\nu(\hat{E}_i) = \sum_j \nu(\hat{E}_{ij}) = \nu(\sum_j \hat{E}_{ij}). \tag{4}$$

ν, which represents the preparation procedure, must of course be independent of $\hat{E}_i$, $\hat{E}_{ij}$. The above implies that $\nu(\hat{E}_i - \sum_j \hat{E}_{ij}) = 0$. The only effect $\hat{A}$ consistent with $\nu(\hat{A}) = 0$ for all ν satisfying P1-P3 is the zero operator. Thus we conclude that

$$\hat{E}_i = \sum_j \hat{E}_{ij}. \tag{5}$$

We will formalise this notion (the physical notion of causality) is formalised in the additional postulate A2 below. In the second measurement, each outcome is represented by an effect $\hat{F}_j$. For each $\hat{F}_j$ and each outcome i of the first measurement, there is a distinct effect $\hat{E}_{ij}$ describing the joint measurement. Thus, for each i we can define a map $\hat{F}_j \to \hat{E}_{ij} = \mathcal{T}_i(\hat{F}_j)$. Further, for each i and for any given measure $\nu(\hat{E}_{ij})$ on the joint measurement procedure, we require that the statistics of the second measurement can be derived from some (sub-normalised) measure over $\hat{F}_j$. That is,

$$\mathrm{P}(i,j) = \mu^i_\nu(\hat{F}_j) \tag{6}$$

where the notation μ^i_ν indicates that the measure depends on both ν and i. To be clear, we assume that ν satisfies propositions P1 - P3 and our additional assumptions on the *joint* measure μ^i_ν are:

(A0) $\mu^i_\nu(\hat{F}_j) = \nu(\hat{E}_{ij}) = \nu(\mathcal{T}_i(\hat{F}_j))$

(A1) $0 \leq \mu^i_\nu(\hat{F}_j) \leq \nu(\hat{E}_i) < 1$.

(A2) $\mu^i_\nu(\hat{I}) = \nu(\hat{E}_i)$.

(A3) $\mu^i_\nu(\hat{F}_j + \hat{F}_k + \ldots) = \mu^i_\nu(\hat{F}_j) + \mu^i_\nu(\hat{F}_k) + \ldots$

These *additional* postulates are analogous to those used by Busch, modified to allow for the idea of conditionality. The probability rule is derived from these alongside our above definition of a measurement as a Hilbert space operator, inherent in which is the subspace structure, and assuming a fixed causal order. We first note that according to the arguments given by Busch [2] we can extend any μ^i_ν satisfying the additivity property (A3) to full linearity on all positive operators:

$$\mu^i_\nu(\alpha\hat{F}_j + \beta\hat{F}_k + \ldots) = \alpha\mu^i_\nu(\hat{F}_j) + \beta\mu^i_\nu(\hat{F}_k) + \ldots \tag{7}$$

where $\alpha, \beta \geq 0$. Thus it follows from proposition (P0′) that

$$\begin{aligned}\nu(\mathcal{T}_i(\alpha\hat{F}_j + \beta\hat{F}_k + \ldots)) &= \alpha\nu(\mathcal{T}_i(\hat{F}_j)) + \beta\nu(\mathcal{T}_i(\hat{F}_k)) + \ldots \\ &= \nu(\alpha\mathcal{T}_i(\hat{F}_j) + \beta\mathcal{T}_i(\hat{F}_k)) + \ldots\end{aligned} \tag{8}$$

where in the last line we have used linearity of ν, which follows from Busch's original proof. Finally, as we require this hold for all ν, we obtain

$$\mathcal{T}_i(\alpha\hat{F}_j + \beta\hat{F}_k + \ldots) = \alpha\mathcal{T}_i(\hat{F}_j) + \beta\mathcal{T}_i(\hat{F}_k)) + \ldots \tag{9}$$

We can readily extend linearity on positive operators to linearity on all Hermitian operators [2, 14, 15], from which we obtain that $\mathcal{T}_i$ is a linear operator on the (real) vector space of Hermitian operators. Thus we find that the most general joint measure $\mathrm{P}(i,j)$ satisfying the propositions (A0) and (A3) is of the form

$$\mathrm{P}(i,j) = \mu^i_\nu(\hat{F}_j) = \mathrm{Tr}\left(\hat{\rho}\mathcal{T}_i\left(\hat{F}_j\right)\right) \tag{10}$$

for some linear transformation $\mathcal{T}_i$. Note that the presence of an intermediate measurement is accomodated mathematically through exactly the non-uniqueness of inner product discussed earlier.

We have not yet addressed propositions (A1) and (A2), and we return to these now. It is perhaps clearest to explicitly write the operators $\hat{\rho}$, $\hat{F}_j$ as vectors in the space of Hermitian operators on Hilbert space. We use Liouville space notation (see e.g. [28, 29]), in which

$$|i\rangle\langle j| \leftrightarrow |ij^\dagger\rangle\rangle. \tag{11}$$

In this notation, any operator $\hat{A} = \sum_{ij} a_{ij}|i\rangle\langle j|$ is therefore represented by a Liouville space vector

$$|A\rangle\rangle = \sum_{ij} a_{ij}|ij^\dagger\rangle\rangle. \tag{12}$$

Further, the inner product $\mathrm{Tr}(\hat{A}^\dagger\hat{B})$ is expressed:

$$\mathrm{Tr}(\hat{A}^\dagger\hat{B}) = \sum_{ij} a^*_{ij}b_{ij} = \langle\langle A|B\rangle\rangle, \tag{13}$$

and thus our probability rule, Eqn. (10) may be written:

$$\mathrm{P}(i,j) = \langle\langle\rho|T_i|F_j\rangle\rangle \tag{14}$$

where T_i is an operator on Liouville space. Denoting the Hilbert space on which $\hat{\rho}$, $\hat{F}_j$ are defined respectively as $\mathcal{H}_{\mathrm{in}}$, $\mathcal{H}_{\mathrm{out}}$, T_i is thus a linear operator

$$T_i : \mathcal{H}_{\mathrm{out}} \otimes \mathcal{H}^\dagger_{\mathrm{out}} \to \mathcal{H}_{\mathrm{in}} \otimes \mathcal{H}^\dagger_{\mathrm{in}}. \tag{15}$$

We use the subscripts "in" and "out" in the remainder of the paper to distinguish between those indices associated with the states $\hat{\rho}$, and those associated with the measurement $\hat{F}_j$ respectively, wherever this is required for clarity.

Alternatively, we can write

$$\mathrm{P}(i,j) = \mathrm{Tr}(T_i|F_j\rangle\rangle\langle\langle\rho|). \tag{16}$$

In the same way as we can consider $\hat{\rho}$ alternatively to be an operator in Hilbert space or a Liouville space vector, it is convenient to consider T_i to be a vector on the space $\mathcal{H}_{\rm in}\otimes\mathcal{H}_{\rm in}^\dagger\otimes\mathcal{H}_{\rm out}\otimes\mathcal{H}_{\rm out}^\dagger$. In what follows, we can then interpret this as an operator on various spaces, as appropriate. We first note that we require this probability to be a real number. Each of $\hat{\rho}$, $\hat{F}_j$ are positive operators, and thus may be expressed as positive linear combinations of pure states. Thus, without loss of generality, we can consider pure states only, $\hat{\rho} = |\psi\rangle\langle\psi|$, $\hat{F}_j = |m_j\rangle\langle m_j|$ and denote

$$\begin{aligned}|F_j\rangle\rangle\langle\langle\rho| &= |m_j m_j^\dagger\rangle\rangle\langle\langle\psi\psi^\dagger| \\ &\leftrightarrow (|\psi\rangle_{\rm in}\otimes|m_j\rangle_{\rm out})(\langle\psi|_{\rm in}\otimes\langle m_j|_{\rm out})\end{aligned} \tag{17}$$

We thus find that if we interpret T_i as an operator on $\mathcal{H}_{\rm in}\otimes\mathcal{H}_{\rm out}$, the requirement that probabilities be real implies that it be a Hermitian operator on this space. Explicitly, define T_i' to be that operator on $\mathcal{H}_{\rm in}\otimes\mathcal{H}_{\rm out}$ such that

$$\begin{aligned}\mathrm{P}(i,j) &= \mathrm{Tr}(T_i|m_j m_j^\dagger\rangle\rangle\langle\langle\psi\psi^\dagger|) \\ &= \mathrm{Tr}(T_i'|\psi\rangle_{\rm in}\otimes|m_j\rangle_{\rm out}\langle\psi|_{\rm in}\otimes\langle m_j|_{\rm out}),\end{aligned} \tag{18}$$

which implies that that the matrix elements satisfy

$$\langle\langle j_{in} i_{in}^\dagger|T_i'|l_{out}k_{out}^\dagger\rangle\rangle = \langle\langle i_{in} j_{in}^\dagger|T_i|k_{out}l_{out}^\dagger\rangle\rangle. \tag{19}$$

We then require T_i' to be a Hermitian operator. This interpretation is precisely the Jamiolkowski form of a map [9, 19], and is seen to arise naturally in this approach.

It remains to impose positivity and normalisation of our joint probability. It is clear from the discussion so far that positivity requires that T_i, when interpreted as above, have a positive expectation value for all product states in $\mathcal{H}_{\rm in}\otimes\mathcal{H}_{\rm out}$. We require also the more stringent constraint of complete positivity: if T_i acts only on a subsystem (A) of a larger system (AB), all probability measures on the joint system must remain positive:

$$\langle\langle\rho_{AB}|T_i^A\otimes I^B|F_{jAB}\rangle\rangle \geq 0. \tag{20}$$

The requirement of complete positivity follows from the structure of effects and the requirement of positivity of probabilities (A1). Specifically, the assumption that all positive operators $\hat{E}\leq\hat{I}$ are allowed effects, along with an assumption that all measures are possible in principle imposes the requirement Eqn. (20).

As before, without loss of generality, we consider $\hat{\rho}$ and $\hat{F}_j$ to be pure states. We further write $|\psi\rangle$ in the Schmidt basis:

$$|\psi\rangle = \sum_i \lambda_i|i\rangle_A|i\rangle_B \tag{21}$$

where $\lambda_i > 0$. In this basis we denote

$$|m_j\rangle = \sum_{ik} c_{ik}^{(j)}|i\rangle_A|k\rangle_B. \tag{22}$$

Thus

$$\begin{aligned}|\rho_{AB}\rangle\rangle &= \sum_{ik}\lambda_i\lambda_k|i_A i_B k_A^\dagger k_B^\dagger\rangle\rangle \\ |F_{jAB}\rangle\rangle &= \sum_{iklm} c_{ik}^{(j)}c_{lm}^{(j)*}|i_A k_B l_A^\dagger m_B^\dagger\rangle\rangle\end{aligned} \tag{23}$$

Putting all this together and simplifying gives:

$$\begin{aligned}&\langle\langle\rho_{AB}|T_i^A\otimes I^B|F_{jAB}\rangle\rangle \\ &= \sum_{iklmnp}\lambda_i\lambda_k c_{lm}^{(j)}c_{np}^{(j)*}\langle\langle i_A i_B k_A^\dagger k_B^\dagger|T_i^A\otimes I^B|l_A m_B n_A^\dagger p_B^\dagger\rangle\rangle \\ &= \sum_{iklmnp}\lambda_i\lambda_k c_{lm}^{(j)}c_{np}^{(j)*}\langle\langle i_A k_A^\dagger|T_i^A|l_A n_A^\dagger\rangle\rangle\langle\langle i_B k_B^\dagger|m_B p_B^\dagger\rangle\rangle \\ &= \sum_{iklmnp}\lambda_i\lambda_k c_{lm}^{(j)}c_{np}^{(j)*}\langle\langle i_A k_A^\dagger|T_i^A|l_A n_A^\dagger\rangle\rangle\delta_{im}\delta_{kp} \\ &= \sum_{ikln}\lambda_i\lambda_k c_{li}^{(j)}c_{nk}^{(j)*}\langle\langle i_{\rm in}k_{\rm in}^\dagger|T_i|l_{\rm out}n_{\rm out}^\dagger\rangle\rangle,\end{aligned} \tag{24}$$

where in the last line we have dropped the label A, which is no longer needed, and introduced subscripts denoting input and output spaces for clarity. The structure of this final line indicates that it would be fruitful to consider the representation of T_i on the space $\mathcal{H}_{\rm in}\otimes\mathcal{H}_{\rm out}^\dagger$. We thus define $\widetilde{T}_i$ to be that operator on this space such that

$$\langle\langle i_{\rm in}l_{\rm out}^\dagger|\widetilde{T}_i|k_{\rm in}n_{\rm out}^\dagger\rangle\rangle = \langle\langle i_{\rm in}k_{\rm in}^\dagger|T_i|l_{\rm out}n_{\rm out}^\dagger\rangle\rangle, \tag{25}$$

where, for convenience, we have extended the concept of Liouville space vector in the natural way to include cases in which the "bra" and "ket" vectors may be on different spaces. Thus we obtain

$$\begin{aligned}&\langle\langle\rho_{AB}|T_i^A\otimes I^B|F_{jAB}\rangle\rangle \\ &= \left(\sum_{il}\lambda_i c_{li}^{(j)}\langle\langle i_{\rm in}l_{\rm out}^\dagger|\right)\widetilde{T}_i\left(\sum_{kn}\lambda_k c_{nk}^{(j)*}|k_{\rm in}n_{\rm out}^\dagger\rangle\rangle\right).\end{aligned} \tag{26}$$

Finally, denoting

$$|\Phi_j\rangle\rangle = \sum_{kn}\lambda_k c_{nk}^{(j)*}|k_{\rm in}n_{\rm out}^\dagger\rangle\rangle \tag{27}$$

we see that our probability rule has the rather compact form

$$\mathrm{P}(i,j) = \langle\langle\rho_{AB}|T_i^A\otimes I^B|F_{jAB}\rangle\rangle = \langle\langle\Phi_j|\widetilde{T}_i|\Phi_j\rangle\rangle. \tag{28}$$

Thus $|\Phi_j\rangle\rangle$ is a state on the space $\mathcal{H}_{\rm in}\otimes\mathcal{H}_{\rm out}^\dagger$, and the discussion above shows that, in general, this need not be a product state. Thus we require that T_i, when interpreted as an operator on this space (i.e. $\widetilde{T}_i$), be a positive operator. This is the Choi form of a map [8, 19], and again, is seen to arise rather naturally in this approach. As $\widetilde{T}_i$ is a positive operator, it has an eigendecomposition

$$\widetilde{T}_i = \sum_k |\alpha_{ik}\rangle\rangle\langle\langle\alpha_{ik}| \tag{29}$$

where $|\alpha_{ik}\rangle\rangle = \sum_{lm} \alpha_{lm}^{(ik)} |l_{\rm in} m_{\rm out}^\dagger\rangle\rangle$ is not normalised. Thus any joint probability satisfying the propositions has the form

$$\begin{aligned} \mathrm{P}(i,j) &= \sum_k |\langle\langle\alpha_{ik}|\Phi_j\rangle\rangle|^2 \\ &= \sum_k \left| \left(\sum_{lm} \alpha_{lm}^{(ik)*} \langle\langle l_{\rm in} m_{\rm out}^\dagger| \right) \right. \\ &\quad \left. \times \left(\sum_{np} \lambda_n c_{pn}^{(j)*} |n_{\rm in} p_{\rm out}^\dagger\rangle\rangle \right) \right|^2 \\ &= \sum_k \left| \sum_{np} \lambda_n c_{pn}^{(j)*} \langle p|_{\rm out} \hat{A}_{ik} |n\rangle_{\rm in} \right|^2, \end{aligned} \tag{30}$$

where $\hat{A}_{ik} = \sum_{lm} \alpha_{lm}^{(ik)*} |m\rangle_{\rm out}\langle l|_{\rm in}$. Note that the combination of preparation and measurement is described by a so-called entangled two-time state [30, 31]. A two time state may be used to describe pre- and post-selection, and is comprised of a state vector describing the preparation, and one describing a later measurement [32–34]; in the language of Hilbert spaces, a vector on $\mathcal{H}_{out} \otimes \widetilde{\mathcal{H}}_{in}^\dagger$. Non-product states arise in exactly the way we have seen here, through pre- and post-selections which are entangled with another system.

For the product state case, in which the coefficients c_{pn} are independent of n, we have $\hat{\rho} = \sum_{ij} \lambda_i \lambda_j |i\rangle\langle j|$, $\hat{F}_j = \sum_{mn} c_m^{(j)} c_n^{(j)*} |m\rangle\langle n|$, and we obtain the familiar sequential measurement rule:

$$\begin{aligned} \mathrm{P}(i,j) &= \sum_k \left| \sum_{np} \lambda_n c_p^{(j)*} \langle p|_{\rm out} \hat{A}_{ik} |n\rangle_{\rm in} \right|^2 \\ &= \mathrm{Tr}\left(\hat{F}_j \sum_k \hat{A}_{ik} \hat{\rho} \hat{A}_{ik}^\dagger \right). \end{aligned} \tag{31}$$

Finally, we return to normalisation of the measure: proposition (P2′) is satisfied if

$$\mathrm{Tr}(\rho \hat{E}_i) = \mathrm{Tr}(\sum_k \hat{A}_{ik} \hat{\rho} \hat{A}_{ik}^\dagger) = \mathrm{Tr}(\hat{\rho} \sum_k \hat{A}_{ik}^\dagger \hat{A}_{ik}). \tag{32}$$

We thus require $\sum_k \hat{A}_{ik}^\dagger \hat{A}_{ik} = \hat{E}_i$.

We thus obtain the usual Kraus form of a map from a simple extension of the Gleason-Busch theorem: given the assumption that measurements are described by effects, that is positive operators $\hat{E} \leq \hat{I}$, along with some reasonable assumptions (A0-A3), every joint probability over sequential measurements is of the form

$$\mathrm{P}(i,j) = \langle\langle\Phi_j|\widetilde{T}_i|\Phi_j\rangle\rangle \tag{33}$$

where $|\Phi\rangle\rangle$ is a two-time vector (defined on the Hilbert space $\mathcal{H}_{\rm in} \otimes \mathcal{H}_{\rm out}^\dagger$) representing the measurement and preparation, and $\widetilde{T}_i$ is a positive operator on this space. Where $|\Phi_j\rangle\rangle$ is a product state, this reduces to the familiar Kraus form

$$\mathrm{P}(i,j) = \mathrm{Tr}\left(\hat{F}_j \sum_k \hat{A}_{ik} \hat{\rho} \hat{A}_{ik}^\dagger \right). \tag{34}$$

It is readily verified that these probabilities sum to one, as desired:

$$\begin{aligned} \sum_{i,j} \mathrm{P}(i,j) &= \mathrm{Tr}\left(\sum_j \hat{F}_j \sum_{i,k} \hat{A}_{ik} \hat{\rho} \hat{A}_{ik}^\dagger \right) \\ &= \mathrm{Tr}\left(\sum_i \hat{\rho} \sum_k \hat{A}_{ik}^\dagger \hat{A}_{ik} \right) = \mathrm{Tr}\left(\sum_i \hat{\rho} \hat{E}_i \right) = 1. \end{aligned} \tag{35}$$

From this we can further derive conditional probabilities:

$$\begin{aligned} \mathrm{P}(j|i) &= \frac{\mathrm{P}(i,j)}{P(i)} \\ &= \frac{\mathrm{Tr}\left(\hat{F}_j \sum_k \hat{A}_{ik} \hat{\rho} \hat{A}_{ik}^\dagger \right)}{\mathrm{Tr}\left(\hat{\rho} \sum_k \hat{A}_{ik}^\dagger \hat{A}_{ik} \right)} \\ &= \mathrm{Tr}\left(\frac{\sum_k \hat{A}_{ik} \hat{\rho} \hat{A}_{ik}^\dagger}{\mathrm{Tr}\left(\hat{\rho} \sum_k \hat{A}_{ik}^\dagger \hat{A}_{ik} \right)} \hat{F}_j \right). \end{aligned} \tag{36}$$

from which we recover the Kraus update rule:

$$\hat{\rho} \to \hat{\rho}_i = \frac{\sum_k \hat{A}_{ik} \hat{\rho} \hat{A}_{ik}^\dagger}{\mathrm{Tr}\left(\hat{\rho} \sum_k \hat{A}_{ik}^\dagger \hat{A}_{ik} \right)}. \tag{37}$$

To summarise, from the assumption that measurements are associated with effects (positive operators $\hat{E} \leq \hat{I}$) along with some reasonable propositions that measures and joint measures should obey, we find that pre- and post-selections are described by two-time states; intermediate measurements are associated with positive operators on the vector space of two-time states, or alternatively with positive Choi states; and that the state update rule is given by the familiar Kraus form. Up to the particular choice of description (Choi-Jamiolkowski isomorphism / Kraus operator form), this is thus the *unique* way to define joint probabilities over sequential measurements in quantum mechanics.

Herein we have considered just two sequential measurements however our result could be easily generalised to longer chains. One would argue for the preparation and first measurement, represented as the vector $\sum_k \hat{A}_{ik} \otimes \hat{A}_{ik}^\dagger |\rho\rangle\rangle$ as representing an individual preparation procedure. The two measurements in the above procedure would then represent the second and third measurements in the new scenario, and then find the expected three-measurement probability rule using the same method.

We note that a key component of our approach is a proof that intermediate measurements are described by transformations of effects, and that these must be linear. Other proofs of the most general form of transformation [7–10, 35] take linearity as an assumption. Indeed strange and seemingly unphysical things become possible if we allow non-linear evolution in quantum mechanics [36–39]. In our approach the requirement of linearity (and indeed, that an intermediate measurement corresponds to a transformation on effects) follows from the requirement that the statistics of sequential measurements be derived from a measure on the second measurement. We note that we did not assume that the intermediate measurement was associated with a transformation, we simply observed that for each i there exists a mapping between any set of operators describing the joint measurement procedure, and that describing the second measurement alone.

III. PHYSICAL MEANING OF THE AXIOMS

The axioms (A0 - A3) may be considered rather abstractly, as desired properties of probability measures, or can be motivated through physical considerations. Following Hardy [14], we suppose that probabilities are measureable in the following sense: if we repeat an experiment a large number of times N, the fraction of runs in which we observe a particular event i tends to a constant $\frac{N_i}{N}$, which we interpret as a probability $p_i = \frac{N_i}{N}$. Additivity then follows rather naturally from counting events $p_i + p_j = \frac{N_i+N_j}{N}$, while clearly $0 \leq N_i \leq N$.

For the sequential measurement case, we have introduced a rather inocuous "zeroth" proposition (P0′), which corresponds to an assumption of non-contextuality at the level of the description of measurement. Non-contextuality means that the value assigned to a physical quantity is independent of the *context* in which that quantity is measured: that is, independent of anything else which may be measured with it [3]. Gleason's theorem is generally taken as proof that a non-contextual hidden variable model reproducing the predictions of quantum theory is not possible [2, 40]. At the level of operators in Gleason's theorem, non-contextuality means that if an effect $\hat{E}$ is a member of two different sets, the probability associated with $\hat{E}$ is independent of which set we are considering. Physically, this means that if $\hat{E}$ represents a measurement outcome in two different measurements, the probability of seeing this outcome is independent of which measurement is actually performed. Non-contextuality in this sense is implicitly assumed in the Gleason-Busch theorem, in the assumption that each measure is a function on $\hat{E}$ (see also [11] for a discussion of non-contextuality in this context).

In the present work, we assume non-contextuality in the mapping from physical measurement apparatus to mathematical description: that is, if a particular measurement outcome may be associated with an operator $\hat{E}$, then for every physical experiment containing the corresponding apparatus, the probability of obtaining this outcome may be expressed as *some* measure $\nu(\hat{E})$. For the first measurement, the assumption of non-contextuality means that the description of measurement is independent of any post-processing, from which we obtain the requirement $\hat{E}_i = \sum_j \hat{E}_{ij}$. For the second measurement, non-contextuality means that every measure is a linear function of the operators $\{\hat{F}_j\}$, leading to our proposition (P0′). In essence, this is what is meant by the assumption that the measurement is described by operators $\{\hat{F}_j\}$, however as this is the key assumption it is worth being rather explicit about the physical meaning.

IV. DISCUSSION

In this work we have shown that the Gleason-Busch theorem is rich enough to contain the structure not only of single measurement statistics, but also of sequential measurements. We note that we do not at any point assume explicitly that intermediate measurements are associated with transformations, or that these transformations be linear, rather this *emerges* as a consequence of the above considerations. We have given a small set of reasonable, physically motivated axioms, from which the structure of sequential measurements follows.

The Gleason-Busch theorem [1, 2] shows that if measurements are described by effects, the Born rule is the most general probability rule allowed. More recently, Shrapnel et al [13] derived the most general frame function on completely positive maps, with a view to understanding recent work on non-fixed causal order. Our work provides a link between the two, starting from a minimal set of axioms to show that the most general sequential measurement rule consistent with these axioms corresponds to a completely positive map.

We note that we have not explicitly assumed convex linearity on preparations: we do not assume anything about the relationship between ν and μ^i_ν. The linearity of the resulting probability rule in both preparation and measurement emerges as a consequence of the axioms. An alternative approach could argue that mixtures of preparations are allowed, and any probability rule should be linear in these. This further has the advantage of treating preparations and measurements symmetrically. Our aim in the present work was however to provide a set of axioms as close as possible in spirit to the Gleason-Busch theorem, and to assume as little as possible about preparations.

We finally note that we have not assumed in our sequential measurement axioms that the joint probability rule is linear in the effects describing the first measurement. Indeed, it turns out that this is not the case: seeking a rule linear in both sets of effects would be much more restrictive. This is, of course, not contrary to the

Gleason-Busch theorem in its original form, which refers only to statistics and says nothing about state update. The choice of causal order dictates that it is the effects describing the second, and final measurement in which the joint probability rule must be linear.

The formulation arrived at herein, similar to that of Silva et al [34], with the explicit role of pre- and post-selection lends itself in particular to calculations of relevance to quantum cryptography, in which post-measurement information is often made available. We explore these applications elsewhere.

We finish with a comment on the importance of an axiomatic approach to quantum communications. Classical cryptosystems are, of course, rather effective. The practical significance of quantum cryptographic protocols as a technological development has been the subject of some debate (see e.g. [41, 42] and references therein). The oft-cited advantage of quantum key distribution, for example, is that security is contingent only on the laws of quantum mechanics being correct, and not on computational assumptions. For skeptics this begs the question: how confident are we in the correctness of quantum mechanics? An axiomatic approach illuminates exactly what assumptions underlie security proofs.

ACKNOWLEDGMENTS

This work was supported by the UK Engineering and Physical Sciences Research Council and by the Royal Society (RP150122). We thank Mark Hillery for helpful comments and suggestions.

[1] A. M. Gleason, Indiana Univ. Math. J. **6**, 885 (1957).
[2] P. Busch, Phys. Rev. Lett. **91**, 120403 (2003).
[3] A. Peres, *Quantum theory: Concepts and Methods* (Kluwer Academic Publishers, 1998).
[4] S. M. Barnett, J. D. Cresser, J. Jeffers, and D. T. Pegg, New J. Phys **16**, 043025 (2014).
[5] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, 1955).
[6] G. Lüders, Ann. Phys. (Leipzig) **15**, 663 (2006).
[7] K. Kraus, *States, Effects and Operations* (Springer, 1983).
[8] M.-D. Choi, Linear Algebra and its Applications **10**, 285 (1975).
[9] A. Jamiolkowski, Rep. Math. Phys. **4**, 275 (1972).
[10] S. Croke, S. M. Barnett, and S. Stenholm, Annals of physics **323**, 893 (2008).
[11] C. M. Caves, C. A. Fuchs, K. M. Manne, and J. M. Renes, Foundations of Physics **34**, 193 (2004).
[12] G. Cassinelli and N. Zanghi, Il Nuovo Cimento **73B**, 237 (1983).
[13] S. Shrapnel, F. Costa, and G. Milburn, "Updating the born rule," (2017), arxiv: 1702.01845v1 [quant-ph].
[14] L. Hardy, "Quantum theory from five reasonable axioms," (2001), v4 used here: available from arXiv:quant-ph/0101012v4.
[15] J. Barrett, Phys. Rev. A **75**, 032304 (2007).
[16] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. A **84**, 012311 (2011).
[17] L. Masanes and M. P. Müller, New J. Phys. **13**, 063001 (2011).
[18] O. Oreshkov, F. Costa, and C. Brukner, Nat. Commun. **3**, 1092 (2012).
[19] M. S. Leifer and R. W. Spekkens, Phys. Rev. A **88**, 052130 (2013).
[20] F. Costa and S. Shrapnel, New J. Phys. **18**, 063032 (2016).
[21] J. A. Allen, J. Barrett, D. C. Horsman, C. M. Lee, and R. W. Spekkens, Phys. Rev. X **7**, 031021 (2017).
[22] C. J. Isham, *Lectures on Groups and Vector Spaces for Physicists* (World Scientific, Singapore, 1989).
[23] C. M. Bender and S. Boettcher, Phys. Rev. Lett. **80**, 5243 (1998).
[24] C. M. Bender, Reports on Progress in Physics **70**, 947 (2007).
[25] A. Mostafazadeh, J. Phys. A: Math. Gen. **36**, 7081 (2003).
[26] A. Mostafazadeh, Int. J. Geom. Methods Mod. Phys. **7**, 1191 (2010).
[27] S. Croke, Phys. Rev. A **91**, 052113 (2015).
[28] J. Fiutak and J. Van Kranendonk, Canadian Journal of Physics **40**, 1085 (1962).
[29] S. M. Barnett and B. J. Dalton, J. Phys. A: Math. Gen. **20**, 411 (1987).
[30] Y. Aharonov, P. G. Bergmann, and J. L. Lebowitz, Phys. Rev. **134**, B1410 (1964).
[31] Y. Aharonov and L. Vaidman, in *Time and Quantum Mechanics, Lect. Notes Phys., Vol 734*, edited by J. G. Muga, R. Sala Mayato, and I. L. Egusquiza (Springer, Berlin, 2008) pp. 399–447.
[32] Y. Aharonov and L. Vaidman, J. Phys. A: Math. Gen. **24**, 2315 (1991).
[33] Y. Aharonov, S. Popescu, J. Tollaksen, and L. Vaidman, Phys. Rev. A **79**, 052110 (2009).
[34] R. Silva, Y. Guryanova, N. Brunner, N. Linden, A. J. Short, and S. Popescu, Phys. Rev. A **89**, 012121 (2014).
[35] C. M. Caves, Journal of Superconductivity **12**, 707 (1999).
[36] A. Peres, Phys. Rev. Lett. **63**, 1114 (1989).
[37] N. Gisin, Phys. Lett. A **143**, 1 (1990).
[38] J. Polchinski, Phys. Rev. Lett. **66**, 397 (1991).
[39] D. S. Abrams and S. Lloyd, Phys. Rev. Lett. **81**, 3992 (1998).
[40] J. S. Bell, Rev. Mod. Phys. **38** (1966), reprinted in Bell, J.S., Speakable and Unspeakable in Quantum Mechanics. Cambridge University Press, Cambridge (1987).
[41] D. Stebila, M. Mosca, and N. Lütkenhaus, in *LNICST 36, Quantum Communication and Quantum Networking*, edited by A. Sergienko, S. Pascazio, and P. Villoresi (Springer, Berlin, 2010) pp. 283–296.
[42] V. Scarani, H. Bechmann-Pasquinucci, N. J. Cerf, M. Dušek, N. Lütkenhaus, and M. Peev, Rev. Mod. Phys. **81**, 1301 (2009).