IOPscience iopscience.iop.org

Home Search Collections Journals About Contact us My IOPscience

# Quantum probability rule: a generalization of the theorems of Gleason and Busch

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2014 New J. Phys. 16 043025

(http://iopscience.iop.org/1367-2630/16/4/043025)

View the table of contents for this issue, or go to the journal homepage for more

Download details:

IP Address: 130.159.82.136
This content was downloaded on 25/11/2014 at 11:06

Please note that terms and conditions apply.

New Journal of Physics
The open access journal at the forefront of physics
Deutsche Physikalische Gesellschaft DPG | IOP Institute of Physics

# Quantum probability rule: a generalization of the theorems of Gleason and Busch

**Stephen M Barnett[1], James D Cresser[2,3], John Jeffers[3] and David T Pegg[4]**
[1] School of Physics and Astronomy, University of Glasgow, Kelvin Building, University Avenue, Glasgow G12 8QQ, UK
[2] Department of Physics and Astronomy, Faculty of Science, Macquarie University, NSW 2109, Australia
[3] Department of Physics, University of Strathclyde, Glasgow G4 0NG, UK
[4] Centre for Quantum Dynamics, Griffith University, Nathan, Brisbane, QLD 4111, Australia
E-mail: john.jeffers@strath.ac.uk

Received 20 January 2014, revised 5 March 2014
Accepted for publication 21 March 2014
Published 29 April 2014
*New Journal of Physics* **16** (2014) 043025
doi:10.1088/1367-2630/16/4/043025

**Abstract**
Busch's theorem deriving the standard quantum probability rule can be regarded as a more general form of Gleason's theorem. Here we show that a further generalization is possible by reducing the number of quantum postulates used by Busch. We do not assume that the positive measurement outcome operators are effects or that they form a probability operator measure. We derive a more general probability rule from which the standard rule can be obtained from the normal laws of probability when there is no measurement outcome information available, without the need for further quantum postulates. Our general probability rule has prediction–retrodiction symmetry and we show how it may be applied in quantum communications and in retrodictive quantum theory.

Keywords: quantum probability, quantum information, quantum theory

## 1. Introduction

In the probabilistic interpretation of quantum measurement we have on one hand the physical process of preparing a system in some state and then performing a measurement procedure with the outcomes recorded, allowing probabilities which depend both on the measurement

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence.
Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.


© 2014 IOP Publishing Ltd and Deutsche Physikalische Gesellschaft

procedure and on the preparation process to be determined from the records of many experiments. On the other hand we have the mathematics of Hilbert space entities. To link the two it is axiomatic that there must be some postulate connecting a Hilbert space entity with something physical. The standard quantum probability rule that does this has been highly successful for predicting the outcomes of measurements. This rule could simply be accepted as the required postulate but it may be possible to obtain a better understanding of quantum theory if the rule could be deduced from more fundamental quantum postulates. Gleason's theorem shows, given reasonable assumptions, that quantum probabilities must be expressible as expectation values of projectors or, more precisely, as the trace of the product of a projector and a density operator [1]. This fundamental theorem is of central importance in quantum theory but although it is discussed in some textbooks [2, 3] a derivation of it rarely appears, doubtless because of the complexity of Gleason's proof.

Busch has provided a remarkable extension of Gleason's theorem [4]. It is remarkable in three ways: (i) it applies to state spaces of any dimension whereas Gleason's proof only applies for dimensions greater than two, (ii) it extends Gleason's proof by including generalized measurements [3, 5–7] as well as projective ones and (iii) it is far simpler than Gleason's original proof [4].

Busch associates an outcome $m$ from a measurement with an effect $\hat{E}$, that is a positive operator less than the identity which can therefore be an element of a probability operator measure (POM), often also referred to as a positive operator-valued measure. He equates the measurement outcome probability $p\left(m|s\right)$ for a system prepared in some state $s$ with the value of a function $v\left(\hat{E}\right)$ which he requires to have the following three properties

$$\text{(P1)} \qquad 0 \leqslant v\left(\hat{E}\right) \leqslant 1 \quad \forall\, \hat{E}$$

$$\text{(P2)} \qquad v\left(\hat{\mathrm{I}}\right) = 1 \quad \hat{\mathrm{I}} = \text{ identity operator}$$

$$\text{(P3)} \qquad v\left(\hat{E} + \hat{F} + \cdots\right) = v\left(\hat{E}\right) + v\left(\hat{F}\right) + \cdots,$$

where $\hat{F}$, $\cdots$ are also effects. The sum of the effects in (P3) must not exceed $\hat{\mathrm{I}}$. It should be emphasized that these properties are familiar in the theory of generalized measurements, but are derived on the basis of quantum theory [5, 6]. Busch's aim was, and indeed ours is, rather different: the intention is to postulate these properties as *axioms* and to *derive* quantum probabilities from them.

It is not difficult to show from the normalization condition for the probabilities of all possible outcomes combined with (P2) and the additivity condition (P3), that the sum of the effects must be the identity, that is, the unit operator. Thus the effects representing all possible outcomes for a system in state $s$ form a POM. Also these conditions are consistent with the probability $p\left(m|s\right)$ given by $v\left(\hat{E}\right)$ being non-contextual in the sense that it has this value independently of the particular POM to which $\hat{E}$ belongs, that is, it is independent of the particular measuring device as long as the outcome is represented by $\hat{E}$. To see this, let $\hat{E}$ belong to two different POMs whose remaining elements are $\hat{F}_1, \hat{F}_2\cdots$ and $\hat{G}_1, \hat{G}_2\cdots$ corresponding to measuring devices $f$ and $g$ respectively. Then from normalization and additivity we have

$$p\left(m\middle|s,f\right) = 1 - \sum_i v\left(\hat{F}_i\right) = 1 - v\left(\sum_i \hat{F}_i\right) \tag{1}$$

with a corresponding expression for $p\left(m\middle|s,\ g\right)$. Because the elements of each POM must sum to the unit operator, $\sum_i \hat{F}_i = \sum_j \hat{G}_j$ so $p\left(m\middle|s,f\right) = p\left(m\middle|s,\ g\right)$.

Busch's property (P1) is a property of probabilities in general. His quantum postulates, that is, those that concern Hilbert space operators, lie in (P2), (P3) and the association of a measurement outcome with an effect operator $\hat{E}$. In this paper we drop (P2) and weaken the effect quantum postulate so that it becomes a positive operator quantum postulate. This means we are *not* assuming that the operators representing the measurement outcomes are elements of a POM, which means that we no longer need to assume they are effects. We do, however, assume they are bounded positive operators and adopt an additivity postulate similar to (P3) but we no longer limit the sum of measurement outcome operators to be $\leqslant \hat{\mathrm{I}}$. We find that it is possible with this reduced number of quantum postulates to derive a probability rule that is more general than the standard rule $\mathrm{Tr}\left(\hat{E}_i\,\hat{\rho}\right)$. Furthermore we find that we can then deduce the standard rule from the general rule by the use of normal probability laws.

## 2. General probability rule

A measurement procedure for the determination of probabilities from a record of many experiments involving preparation and measurement will include a chosen measurement device and the method for recording the results obtained from it. For example, two measurement events, such as a zero and a one photocount event, might be recorded as separate events or as a single event described as less than two photocounts. As another example, some experiments might not be recorded because of a post-selection procedure, whereby an experiment is ignored in the event of a particular measurement outcome. We also include in the measurement procedure any means by which information can be obtained that affects the possibility of a recorded event. This can include posterior knowledge. For example if it is known that a photodetector will be damaged if subjected to more than a certain number of photons, then an undamaged detector after the detection event will eliminate the possibility of a recording of a larger number of photons. For our purposes here it is sufficient to specify a measurement procedure $x$ mathematically by the set of possible recorded measurement events $\left\{m_1,\ m_2,\ \cdots\right\}$ that can be obtained from it. We shall not be assuming non-contextuality with respect to the measurement procedure $x$ of the probability that a recorded measurement event is $m_i$, so we shall write this probability as $p\left(m_i\middle|s,\ x\right)$ to show that it may depend on $x$ as well as the state $s$.

Our first postulate is that, for a given measurement procedure $x$, each possible recorded event $m_i$ can be associated with a positive bounded Hilbert space[5] operator $\hat{M}_i$, in such a way that $p\left(m_i\middle|s,\ x\right)$ is proportional to some function $u\left(\hat{M}_i\right)$ of this operator, that is,

[5] It suffices, for our purpose, to consider only state spaces of finite dimension. In this way we avoid complications such as observables with continuous spectra. We may incorporate such observables by means of a suitable limiting process, but such considerations would detract from the essential simplicity of the point that we are trying to make.

$$p\left(m_i \middle| s, x\right) = Q(s, x)u\left(\hat{M}_i\right) \tag{2}$$

where the proportionality factor $Q(s, x)$ is the same for all $\hat{M}_i$ of the set of operators $\left\{\hat{M}_1, \hat{M}_2, \cdots\right\}$, which we can now use to specify the measurement procedure $x$. We are not assuming that $Q(s, x)$ is independent of the measurement procedure itself or of the particular state $s$. We note that any set of positive bounded operators $\hat{M}_i$, to which we refer as measurement operators, can define mathematically a measurement procedure and that, while some measurement procedures have reasonably straightforward physical realizations, others may not.

The function $u\left(\hat{M}_i\right)$ may in general be a complex number $\exp\left(i\theta_i\right)w\left(\hat{M}_i\right)$, say, where $w\left(\hat{M}_i\right)$ is a positive number. The positivity of $p\left(m_i \middle| s, x\right)$ for all $m_i$ then requires $Q(s, x)\exp\left(i\theta_i\right)$ to be positive for all $\theta_i$, which in turn requires $\theta_i$ all to have the same value which we write as $\theta$. Thus we can, from (2), write our first postulate in the form

$$p\left(m_i \middle| s, x\right) = N(s, x)w\left(\hat{M}_i\right) \tag{3}$$

where $N(s, x)$ is the positive normalization factor $Q(s, x)\exp\left(i\theta\right)$.

Our second postulate is that the positive function $w\left(\hat{A}\right)$ of any positive bounded operator is additive, that is,

$$w\left(\hat{A} + \hat{B} + \cdots\right) = w\left(\hat{A}\right) + w\left(\hat{B}\right) + \cdots \tag{4}$$

for all positive bounded operators $\hat{A}, \hat{B}, \cdots$.

We use a method similar to that used by Busch to show firstly that this additivity postulate implies linearity with respect to non-negative rational numbers. From additivity we have, for positive integers $r$ and $n$,

$$\begin{aligned} &nw\left(\frac{r\hat{A}}{n}\right) = w\left(r\hat{A}\right) = rw\left(\hat{A}\right) \\ &\Rightarrow \frac{r}{n}w\left(\hat{A}\right) = w\left(\frac{r}{n}\hat{A}\right). \end{aligned} \tag{5}$$

We can then use the additivity and positivity of $w\left(\hat{A}\right)$ in a limiting argument similar to that used by Busch who showed that $\alpha v(\hat{E}) = v\left(\alpha\hat{E}\right)$, where $\alpha$ is real and $0 \leqslant \alpha \leqslant 1$. In our case we find that $\alpha w\left(\hat{A}\right) = w\left(\alpha\,\hat{A}\right)$ where $\alpha$ is any non-negative real number. Combining this result with additivity we obtain the linearity relation

$$w\left(\sum_i \alpha_i\hat{M}_i\right) = \sum_i \alpha_i w\left(\hat{M}_i\right). \tag{6}$$

We are now in a position to prove our first main result. The measurement operator $\hat{M}_i$ is a positive operator so we can write it in the diagonal form:

$$\hat{M}_i = \sum_\ell \lambda_\ell^i \left|\lambda_\ell^i\right\rangle \left\langle \lambda_\ell^i\right|, \tag{7}$$

where $\left\{\left|\lambda_\ell^i\right\rangle\right\}$ are the eigenstates of $\hat{M}_i$ and $\lambda_\ell^i = \mathrm{Tr}\left(\hat{M}_i \left|\lambda_\ell^i\right\rangle \left\langle \lambda_\ell^i\right|\right) \geqslant 0$ are the corresponding eigenvalues, which are all positive. We should note that the positive operators $\left\{\hat{M}_i\right\}$ will, in general, be non-commuting and therefore will have distinct eigenvectors. It follows, using our linearity condition (6) that

$$w\left(\hat{M}_i\right) = \sum_\ell \mathrm{Tr}\left(\hat{M}_i \left|\lambda_\ell^i\right\rangle \left\langle \lambda_\ell^i\right|\right) w\left(\left|\lambda_\ell^i\right\rangle \left\langle \lambda_\ell^i\right|\right). \tag{8}$$

The $w\left(\left|\lambda_\ell^i\right\rangle \left\langle \lambda_\ell^i\right|\right)$ are simply positive numbers, however, and hence we can write

$$w\left(\hat{M}_i\right) = \sum_\ell \mathrm{Tr}\left[\hat{M}_i \left|\lambda_\ell^i\right\rangle \left\langle \lambda_\ell^i\right| w\left(\left|\lambda_\ell^i\right\rangle \left\langle \lambda_\ell^i\right|\right)\right] = \mathrm{Tr}\left(\hat{M}_i \hat{R}_i\right), \tag{9}$$

where $\hat{R}_i$ is a positive operator, the diagonal elements of which, in the $\left\{\lambda_\ell^i\right\}$ basis, are $w\left(\left|\lambda_\ell^i\right\rangle \left\langle \lambda_\ell^i\right|\right)$. Equation (9) gives no such information about the off-diagonal elements of $\hat{R}_i$ so this operator is not completely determined by equation (9) but we can exploit the linearity relation (6) to show that $\hat{R}_i$ must be independent of $\hat{M}_i$ as follows. Linearity and equation (9) require that $w\left(\hat{M}_1\right) + w\left(\hat{M}_2\right)$ equals $\mathrm{Tr}\left[\left(\hat{M}_1 + \hat{M}_2\right)\hat{R}_{12}\right]$ so we must be able to write $w\left(\hat{M}_1\right)$ and $w\left(\hat{M}_2\right)$ in the form $\mathrm{Tr}\left(\hat{M}_1 \hat{R}_{12}\right)$ and $\mathrm{Tr}\left(\hat{M}_2 \hat{R}_{12}\right)$ respectively, where the common operator $\hat{R}_{12}$ has diagonal elements $w\left(\left|\lambda_\ell^1\right\rangle \left\langle \lambda_\ell^1\right|\right)$ in the $\left\{\lambda_\ell^1\right\}$ basis and $w\left(\left|\lambda_\ell^2\right\rangle \left\langle \lambda_\ell^2\right|\right)$ in the $\left\{\lambda_\ell^2\right\}$ basis. We can combine $\hat{M}_1$ with *any* other positive bounded operators to form a set defining a measurement procedure so the common operator $\hat{R}$ must have diagonal elements $w\left(\left|\lambda_\ell\right\rangle \left\langle \lambda_\ell\right|\right)$ in *any* basis $\left\{\lambda_\ell\right\}$ and thus is independent of any particular $\hat{M}_i$. We can then write

$$w\left(\hat{M}_i\right) = \mathrm{Tr}\left(\hat{M}_i \hat{R}\right) \tag{10}$$

for all $\hat{M}_i$, showing that the probability that a measurement event is $m_i$ depends both on the associated measurement operator and an independent operator, which it is natural to associate physically with the preparation process. We can show that the common operator $\hat{R}$ is unique by using the lemma that two operators having the same diagonal elements in all bases must be equal. We prove this lemma in the appendix.

To obtain the probability $p\left(m_i|s, x\right)$ we require the proportionality factor $N(s, x)$, which can be found from the normalization condition that the probabilities of all possible outcomes sum to unity. This yields

$$p\left(m_i|s, x\right) = \frac{\mathrm{Tr}\left(\hat{M}_i \hat{R}\right)}{\mathrm{Tr}\left(\hat{X}\hat{R}\right)} \tag{11}$$

where $\hat{X} = \sum_j M_j$. Dividing the numerator and denominator by $\mathrm{Tr}\left(\hat{R}\right)$ yields our general probability law

$$p\left(m_i|s,\,x\right) = \frac{\mathrm{Tr}\left(\hat{M}_i\,\hat{\rho}\right)}{\mathrm{Tr}\left(\hat{X}\hat{\rho}\right)}. \tag{12}$$

We note that $\hat{X}$ depends only on the possible recorded measurement outcomes, thereby characterizing the particular measurement procedure $x$, leaving the unit-trace positive $\hat{\rho}$ as a density operator to characterize the state $s$. This is the first main result of the paper: if we reduce the number of Busch's quantum postulates by discarding (P2) and, relaxing the assumption that the operator representing a measurement outcome must be an effect to simply being a positive bounded operator, we arrive at a probability law that any set of positive operators (with finite eigenvalues) can provide a set of probabilities and that these probabilities are calculated using (12).

Before proceeding, we give a simple illustration of the meaning of our second postulate, the additivity postulate. The measurement procedure $x$ only enters into (12) as the sum $\hat{X}$. Consider a particular measuring device with, among other measurement events $m_3$, $m_4$, $\cdots$, the events $m_1$ and $m_2$ corresponding to $\hat{M}_1$ and $\hat{M}_2$ if these are recorded separately. If we record these events together as one event $m_1$ *or* $m_2$ our additivity postulate implies that the corresponding measurement operator is $\hat{M}_1 + \hat{M}_2$. The sum $\hat{X}$ is thus the same whether the measurement procedure involves separately recorded events or a single combined event. As a result of this, while $p\left(m_3|s,\,x\right)$ depends on whether $m_4$ is a possible recorded event or not, it does not depend on whether $m_1$ and $m_2$ are recorded together or separately.

## 3. Standard probability rule

It remains for us to determine the physical meaning of our general probability law. In doing so we arrive, very naturally, at a Bayesian interpretation. Consider the case where we know that a number of possible states $s_k$, for which the density operators are $\hat{\rho}_k$, have probabilities $p_k$ of being the prepared state. The state $s$ based on this knowledge will have a density operator $\hat{\rho} = \sum_k p_k \hat{\rho}_k$ representing the average or *a priori* density operator and the probability of the recorded event being $m_i$ will be given by (12). If the state actually prepared was $s_k$, say, then in place of (12) we would have a different probability

$$p\left(m_i|s_k,\,x\right) = \frac{\mathrm{Tr}\left(\hat{M}_i\,\hat{\rho}_k\right)}{\mathrm{Tr}\left(\hat{X}\hat{\rho}_k\right)}. \tag{13}$$

We should be able to obtain (12) as a sum of these objects, suitably weighted by a probability:

$$\begin{aligned} p\left(m_i|s,\,x\right) &= \sum_k p\left(m_i|s_k,\,x\right)P_k \\ \Rightarrow \mathrm{Tr}\left(\hat{M}_i\,\hat{\rho}\right) &= \sum_k \mathrm{Tr}\left(\hat{M}_i\,\hat{\rho}_k\right)P_k\frac{\mathrm{Tr}\left(\hat{X}\hat{\rho}\right)}{\mathrm{Tr}\left(\hat{X}\hat{\rho}_k\right)} \\ &= \sum_k \mathrm{Tr}\left(\hat{M}_i\,\hat{\rho}_k\right)p_k. \end{aligned} \tag{14}$$

For this to hold in general we need only to set

$$P_k = \frac{\mathrm{Tr}\left(\hat{X}\hat{\rho}_k\right)}{\mathrm{Tr}\left(\hat{X}\hat{\rho}\right)} p_k. \tag{15}$$

The fact that both the $P_k$ and the $p_k$ are probabilities means that their ratio is a likelihood [8], which we can interpret as the likelihood of $s_k$ given $x$:

$$l\left(s_k|x\right) = \frac{\mathrm{Tr}\left(\hat{X}\hat{\rho}_k\right)}{\mathrm{Tr}\left(\hat{X}\hat{\rho}\right)}. \tag{16}$$

In order to adopt this interpretation it is necessary to interpret $P_k$ as an *a posteriori* probability based on some knowledge relating to the recorded outcome of the measurement. Specifically this will be knowledge affecting the possibility that some outcomes may occur. As $\hat{X}$ is the sum of the operators representing the possible recorded measurement outcomes, its value will depend on this knowledge. Then $P_k$ will also depend on this knowledge from equation (15). The simplest example is where the actual outcome itself is known to be $m_i$, say, and then $\hat{X} = \hat{M}_i$ as no other outcomes are possible any longer. We then find from equation (12) that the *a posteriori* probability that the outcome is $m_i$ is unity as it must be. Another example is where joint events $(s, m)$ showing the input state and the consequent measurement outcome are recorded after a known post-selection procedure has rejected some joint events containing particular measurement outcomes. This has the effect of reducing the number of possible recorded outcomes and thus the sum of the operators representing them. In this context $P_k$ is just the probability that the state in a recorded joint event is $s_k$.

We shall express the *a posteriori* nature of $P_k$ by writing it as $P\left(s_k|x\right)$, that is, the probability that state $s_k$ was prepared in a recorded experiment conditioned on the operator corresponding to the measurement outcome being limited to one of the reduced number of terms in the posterior expression for $\hat{X}$. This leads us in turn to interpret $p\left(m_i|s, x\right)$ in equation (12) as

$$\begin{aligned} p\left(m_i|s, x\right) &= \frac{\mathrm{Tr}\left(\hat{M}_i\hat{\rho}\right)}{\mathrm{Tr}\left(\hat{X}\hat{\rho}\right)} = \sum_k \frac{\mathrm{Tr}\left(\hat{M}_i\hat{\rho}_k\right)}{\mathrm{Tr}\left(\hat{X}\hat{\rho}_k\right)} P\left(s_k|x\right) \\ &= \sum_k p\left(m_i|s_k, x\right) P\left(s_k|x\right), \end{aligned} \tag{17}$$

which is consistent with Bayesian probability, confirming our interpretation of $P\left(s_k|x\right)$.

If there is no post-selection and no posterior knowledge about measurement results that can eliminate or reduce the possibility of particular measurement events and thus of the preparation events that may have produced them, then the *a posteriori* probability $P\left(s_k|x\right)$ that any state $s_k$ has occurred must be equal to the *a priori* probability $p_k$ that this state occurs. In this case we have, from equation (15)

$$\mathrm{Tr}\left(\hat{X}\hat{\rho}_k\right) = \mathrm{Tr}\left(\hat{X}\hat{\rho}\right) \tag{18}$$

for all $\hat{\rho}_k$. Consider two density operators $\hat{\rho}_k$ with $k = 1, 2$ related by a unitary transformation $\hat{\rho}_2 = \hat{U}\hat{\rho}_1\hat{U}^{-1}$. From equation (18) we then have

$$\mathrm{Tr}\left(\hat{U}\hat{X}\hat{\rho}_1\hat{U}^{-1}\right) = \mathrm{Tr}\left(\hat{X}\hat{\rho}_1\right) = \mathrm{Tr}\left(\hat{X}\hat{\rho}_2\right) = \mathrm{Tr}\left(\hat{X}\hat{U}\hat{\rho}_1\hat{U}^{-1}\right). \tag{19}$$

For this to hold for any $\hat{\rho}_1$, $\hat{X}$ must commute with any $\hat{U}$ and must therefore be proportional to the unit operator, that is $\hat{X} = K\hat{\mathrm{I}}$. Then our general probability rule (12) becomes the standard, or restricted, probability law

$$p\left(m_i\middle|s\right) = \mathrm{Tr}\left(K^{-1}\hat{M}_i\hat{\rho}\right) = \mathrm{Tr}\left(\hat{E}_i\hat{\rho}\right) \tag{20}$$

say, where

$$\sum_i \hat{E}_i = K^{-1}\sum_i \hat{M}_i = K^{-1}\hat{X} = \hat{\mathrm{I}}. \tag{21}$$

$\hat{E}_i$ are therefore effects and form a POM. In Busch's notation $p\left(m_i\middle|s\right)$ equals $v\left(\hat{E}_i\right)$. Using the latter expression for the left side of equation (20) and then summing both sides over $i$ gives condition (P2), which we see is a *result* of our approach, obtained from our general formula (12) by the usual rules of probability, rather than being an additional quantum postulate.

## 4. Applications

It is natural to ask whether there are any applications of our more general probability formula (12). Here we present three such applications. An obvious, but often overlooked, one is to measurement probabilities when we have some (incomplete) information about the measurement outcome. It is often the case in quantum optics experiments, for example, that we restrict our attention to probabilities given some future event, such as a two-photon cascade in which the detection of one photon is used to herald the emission of another [9]. In such cases $\hat{X}$ will be restricted to only those measurement event operators $\hat{M}_i$ that include the heralding event.

A second example arises in the theory of quantum communications [7]. Here a transmitting party, Alice, selects from a set of possible states $s_i$, with density operators $\hat{\rho}_i$ and prior selection probabilities $p_i$, and sends a quantum system prepared in this state to a receiving party, Bob. Bob's task is to determine from a measurement, as well as possible, the state prepared by Alice. As he knows from the measurement that the outcome is $m_j$ corresponding to $\hat{M}_j$, say, he knows that $\hat{X}$ contains just this single term, that is, his knowledge has eliminated the possibility of all other terms. He can therefore simply write the sum of the possible terms as $\hat{X} = \hat{M}_j$ and obtain from (15) the *a posteriori*, or retrodictive, probability that Alice sent the system in state $s_k$

$$P\left(s_k\middle|m_j\right) = \frac{\mathrm{Tr}\left(\hat{M}_j\hat{\rho}_k p_k\right)}{\mathrm{Tr}\left(\hat{M}_j\hat{\rho}\right)}. \tag{22}$$

We note that retrodictive probabilities such as this can also be found by using Bayes' theorem in conjunction with the usual expression for the quantum probability $\mathrm{Tr}\left(\hat{E}_j\hat{\rho}\right)$ [10]. Peres [11] has described an expression equivalent to equation (22) as the only retrodictive form that can be legitimately derived from conventional quantum mechanics. However here there is no need to add a Bayes rule; it is already contained in the general probability law (12) expressed in the form (15). We note that there is symmetry between the retrodictive form of our probability law (22) and the predictive form (12) which we write here as

$$p\left(m_k\middle|s_j,\,x\right)=\frac{\mathrm{Tr}\left(\hat{\rho}_j\hat{M}_k\right)}{\mathrm{Tr}\left(\hat{\rho}_j\hat{X}\right)} \tag{23}$$

with $\hat{\rho}_k p_k$ in (22) corresponding to $\hat{M}_k$ in (23), $\hat{M}_j$ in (22) corresponding to $\hat{\rho}_j$ in (23) and thus $\hat{\rho}$ in (22) corresponding to $\hat{X}=\sum_k\hat{M}_k$ in (23). This allows Bob an alternative and equivalent way to retrodict by defining a density operator $\hat{\rho}_j$ for a 'retrodictive state' as $\hat{M}_j\big/\mathrm{Tr}\left(\hat{M}_j\right)$, writing $\hat{M}_k$ as $\hat{\rho}_k p_k$ and writing $\hat{X}$ as $\hat{\rho}$ and then substituting into the right side of the predictive formula (23) to obtain the retrodictive expression (22). In this way the general probability rule (12) can be used for *both* prediction and retrodiction without the need to invoke Bayes' theorem, which is already effectively contained in the law. If there is a time interval between preparation and measurement, then Alice would need to allow for evolution of her predictive state in this interval to calculate the probability of a measurement event and Bob would need to allow for the retroevolution of the retrodictive state to retrodict a preparation event.

Our final example completes the resolution of a long-standing controversy in retrodictive quantum theory [12]. In retrodictive quantum theory we assign a retroevolving quantum state on the basis of a *later* measurement and can use this to ask questions about, among other things, initial preparation events. It has been suggested that we can only apply quantum retrodiction if there is no prior information about the preparation event so the prior initial density operator has an unbiased form and is proportional to the identity operator [13, 14]. This is a result of attempting to find a retrodictive formula by making the restricted predictive probability $\mathrm{Tr}\left(\hat{E}_i\hat{\rho}\right)$ symmetric or causally neutral [15] or by using a time-reversed form of Gleason's theorem [14]. From the symmetry inherent in our general probability rule, which reduces to the restricted predictive form when $\hat{X}\propto\hat{\mathrm{I}}$, it is easy to see from the correspondence between $\hat{X}$ and $\hat{\rho}$ above that our general retrodictive formula will reduce to the restricted retrodictive form when $\hat{\rho}\propto\hat{\mathrm{I}}$. To obtain the general, and far more useful, retrodictive probability formula from causal neutrality of a predictive formula it is necessary to start with the general predictive form. For this reason it is also inadequate to use a time-reversed form of Gleason's theorem. Looked at from another view point, retrodicted preparation probabilities are quite often contextual, depending on what other states could possibly be prepared. For example if photon number states are being prepared, there is some limit set by the amount of

energy available or simply by the difficulty in preparing some states[6]. Thus time-reversed theorems incorporating non-contextuality are inappropriate for a general treatment.

## 5. Conclusion

We should note that it is also possible to derive a relationship between Bayes' theorem, predictive and retrodictive quantum theory based on an *assumed* expression for measurement and preparation probabilities in which preparation and measurement operators appear symmetrically [17, 18]. Our general probability rule as derived in this paper, however, enables us to arrive at the correct expression for retrodictive probabilities *without* postulating a symmetric form for the probabilities and thus may be regarded as a more fundamental approach that formally justifies this earlier work.

Busch has relaxed Gleason's postulate that measurement outcomes must be represented by projectors by allowing measurement outcomes to be represented by effects. In this paper we have further relaxed this to the postulate that the probability of a measurement outcome for a particular input state and measurement procedure is proportional to a positive additive function of a bounded positive operator. By allowing the proportionality constant to depend not just on the state but also on the measurement procedure, including choice of measurement device, we are explicitly not assuming non-contextuality in relation to other possible measurement outcomes. Any set of positive operators (with strictly finite eigenvalues) can represent measurement outcomes and can be used to calculate the probabilities of these outcomes. The usually adopted requirement that these operators must sum to the identity is not assumed but *follows* from our approach for the case when there is no prior information about the measurement outcome. This resulting standard, or restricted, probability formula is seen to be a special case of a more general causally-neutral symmetric formula, which can be used for both prediction, that is finding probabilities of measurement outcomes, and for retrodiction involving finding the probabilities of preparation events. When used for prediction, the formula is applicable even when there is partial knowledge of possible measurement outcomes as may occur when post-selection is involved or when there is simply incomplete reporting of outcomes that have occurred. Retrodictive probabilities can, of course, be calculated from the usual restricted formula by employing Bayes' theorem but there is no need to invoke Bayes' theorem when using the general formula. Important examples of the use of our general formula include quantum communications, retrodictive quantum theory and where there is prior agreed post-selection of measurement results.

## Acknowledgments

This work was supported by the Royal Society, Wolfson Foundation and the UK EPSRC. DTP thanks D Matthew Pegg for his support.

[6] This type of situation is considered by Dressel and Jordan [16], who use a symmetric formulation of quantum theory based on quantum instruments (basically corresponding to measurement devices) to derive predictive, retrodictive and 'interdictive' states.

## Appendix. Uniqueness lemma

Here we seek to prove that two operators with the same diagonal elements in all bases must be equal.

Assume that we have two operators $\hat{R}$ and $\hat{Q}$ with the same diagonal elements in any basis. If this is true then for any pair of basis states $|i\rangle$ and $|j\rangle$ we have

$$\langle i|\hat{R}|i\rangle = \langle i|\hat{Q}|i\rangle, \tag{A.1}$$

$$\langle j|\hat{R}|j\rangle = \langle j|\hat{Q}|j\rangle. \tag{A.2}$$

If the diagonal elements are the same in any basis then, for a general superposition of these states

$$|u\rangle = a|i\rangle + b|j\rangle, \tag{A.3}$$

with $a$ and $b$ being any pair of complex amplitudes, we must now also have

$$\langle u|\hat{R}|u\rangle = \langle u|\hat{Q}|u\rangle. \tag{A.4}$$

If this is true for all $a$ and $b$ then it must also be true for the coefficients of $ab^*$ and $a^*b$ in this expression so that

$$\langle i|\hat{R}|j\rangle = \langle i|\hat{Q}|j\rangle, \tag{A.5}$$

$$\langle j|\hat{R}|i\rangle = \langle j|\hat{Q}|i\rangle. \tag{A.6}$$

There is nothing special about the states we have chosen and hence we infer that all the matrix elements of the operators are equal in this basis. If all the matrix elements are equal then the operators must be identical.

## References


[1] Gleason A M 1957 *J. Math. Mech.* **6** 885
[2] Peres A 1995 *Quantum Theory: Concepts and Methods* (Dordrecht: Kluwer)
[3] Isham C J 1995 *Lectures on Quantum Theory* (London: Imperial College Press)
[4] Busch P 2003 *Phys. Rev. Lett.* **91** 120403
[5] Helstrom C W 1976 *Quantum Detection and Estimation Theory* (New York: Academic)
[6] Holevo A S 1982 *Probabilistic and Statistical Aspects of Quantum Theory* 1st edn (Amsterdam: North-Holland)
Holevo A S 2011 *Probabilistic and Statistical Aspects of Quantum Theory* 2nd edn (Pisa: Edizioni della Normale)
[7] Barnett S M 2009 *Quantum Information* (Oxford: Oxford University Press)
[8] Box G E P and Tiao G C 1973 *Bayesian Inference in Statistical Analysis* (New York: Wiley)
[9] Grangier P, Roger G and Aspect A 1986 *Europhys. Lett.* **1** 173
[10] Barnett S M, Pegg D T and Jeffers J 2000 *J. Mod. Opt.* **47** 1779
[11] Peres A 1995 *Phys. Lett.* A **203** 105
[12] Watanabe S 1955 *Rev. Mod. Phys.* **27** 179
Aharonov Y, Bergman P G and Lebowitz J L 1964 *Phys. Rev.* **134** B1410
[13] Belinfante F J 1975 *Measurements and Time Reversal in Objective Quantum Theory* (Oxford: Pergamon)
[14] Amri T, Laurat J and Fabre C 2011 *Phys. Rev. Lett.* **106** 020502

[15] Leifer M S and Spekkens R W 2013 *Phys. Rev.* A **88** 052130 and references therein
[16] Dressel J and Jordan A N 2013 *Phys. Rev.* A **88** 022107
[17] Pegg D T, Barnett S M and Jeffers J 2002 *J. Mod. Opt.* **49** 913
[18] Pegg D T, Barnett S M and Jeffers J 2002 *Phys. Rev.* A **66** 022106