# Gleason-Busch theorem and Bayesian quantum theory 

Barnett, Cresser, Jeffers and Pegg

August 5, 2013-xArqiv

## Abstract

We show that Gleason's theorem, in the form recently generalized by Busch, may be further simplified by dropping one of the three properties from which it was derived. The result is a more general probability than that usually employed in quantum theory in that it shows that any set of positive operators can represent the probabilities for a set of possible events. Remarkably, our more general form seems to contain Bayes's rule for conditional probabilities so there is no need to add it as an additional element. There is no need, moreover, to postulate that the measurement operators sum to the identity; rather this condition follows from our more general rule when there is no prior measurement outcome information available. We show how the new and general probability law may be applied in quantum communications and in retrodictive quantum theory.

Gleason's theorem shows, given reasonable assumptions, that quantum probabilities must be expressible as the 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 text books [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 [5] (ii) it extends Gleason's proof by including generalized measurements $[2,6$, 7] as well as projective ones and (iii) it is far simpler than Gleason's original proof. Let us begin by sketching, very briefly, Busch's proof [4]. Busch begins by introducing a set of probabilities $v(E)$, each of which is obtained from a positive operator $E$. The functions are required to have three properties [8]:
(P1) $0 \leq v(E) \leq 1 \quad \forall E$
(P2) $v(I)=1 \quad I=$ identity operator
(P3) $v(E+F+\cdots)=v(E)+v(F)+\cdots$
Busch combines these to establish a linear dependence on $E$ of $v(E)$ and thence, using a result due to von Neumann, the desired result that the probabilities, $v(E)$, are of the form $\operatorname{Tr}(E \rho)$, where $\rho$ is gthe density operator. Hence we have a set of positive operators $\left\{E_{i}\right\}$, which sum to the identity
and for which the probabilities for the associated measurement outcomes are $\operatorname{Tr}\left(E_{i} \rho\right)$. This corresponds to a probability operator measure (POM) $[6,7]$ or, if you prefer, a positive operator-valued measure (POVM) [2].

In this paper we work from Busch's properties (P1) and (P3) but avoid using (P2). This means that we are not assuming that the $E_{i}$ are elements of a POM. We find that it is still possible to derive a probability rule but that this is more general than $\operatorname{Tr}\left(E_{i} \rho\right)$. We first note that because the $v\left(E_{i}\right)$ are a set of probabilities, it must follow that

$$
\begin{equation*}
\sum_{i} v\left(E_{i}\right)=1 \tag{1}
\end{equation*}
$$

where the sum extends over the positive operators corresponding to the complete set of possible events.

We follow Busch by deriving, on the basis of an additional and reasonable assumption of continuity, the linear dependence of the probabilities:

$$
\begin{equation*}
v\left(\sum_{i=1}^{k} \alpha_{i} E_{i}\right)=\sum_{i=1}^{k} \alpha_{i} v\left(E_{i}\right) \tag{2}
\end{equation*}
$$

where the $\alpha_{i}$ are positive constants so that the $\alpha_{i} E_{i}$ are positive operators. To prove this we introduce a pair of integers $m$ and $n$ and use the property (P3) to show that

$$
\begin{align*}
& v\left(\frac{m}{n} E\right)=m v\left(\frac{1}{n} E\right) \\
& \Rightarrow v\left(\frac{1}{n} E\right)=\left(\frac{1}{m} v\left(\frac{m}{n} E\right)\right. \\
& \Rightarrow \frac{m}{n} v(E)=v\left(\frac{m}{n} E\right) \tag{3}
\end{align*}
$$

If we further require continuity, so that any positive number can be approximated by a nearby rational, then we find $v(\alpha E)=\alpha v(E)$ from which we obtain, using (P3), the required linearity (2).

We are now in a position to prove our main result. The operator $E_{i}$ is a positive operator and it follows that we can write it in the diagonal form:

$$
\begin{equation*}
E_{i}=\sum_{\ell} \lambda_{\ell}^{i}\left|\lambda_{\ell}^{i}\right\rangle\left\langle\lambda_{\ell}^{i}\right| \tag{4}
\end{equation*}
$$

where $\left\{\left|\lambda_{\ell}^{i}\right\rangle\right\}$ are the eigenstates of $E_{i}$ and $\lambda_{\ell}^{i}=\operatorname{Tr}\left(E_{i}\left|\lambda_{\ell}^{i}\right\rangle\left\langle\lambda_{\ell}^{i}\right|\right) \geq 0$ are the corresponding eigenvalues. We should note that the positive operators $\left.\{E)_{i}\right\}$ will, in general, be non-commuting operators and therefore will have distinct eigenvectors. It follows, using our linearity condition (2) that

$$
\begin{equation*}
v\left(E_{i}\right)=\sum_{\ell} \operatorname{Tr}\left(E_{i}\left|\lambda_{\ell}^{i}\right\rangle\left\langle\lambda_{\ell}^{i}\right|\right) v\left(\left|\lambda_{\ell}^{i}\right\rangle\left\langle\lambda_{\ell}^{i}\right|\right) \tag{5}
\end{equation*}
$$

The $v\left(\left|\lambda_{\ell}^{i}\right\rangle\left\langle\lambda_{\ell}^{i}\right|\right)$ are simply positive numbers, however, and hence we can again use linearity to write our probability function in the form

$$
\begin{equation*}
\left.v\left(E_{i}\right)=\sum_{\ell} \operatorname{Tr}\left[E_{i}\left|\lambda_{\ell}^{i}\right\rangle\left\langle\lambda_{\ell}^{i}\right|\right) v\left(\left|\lambda_{\ell}^{i}\right\rangle\left\langle\lambda_{\ell}^{i}\right|\right)\right]=\operatorname{Tr}\left(E_{i} R_{i}\right) \tag{6}
\end{equation*}
$$

where $R_{i}$ is a positive operator, the diagonal elements of which, in the $\left\{\lambda_{\ell}^{i}\right\}$ basis are $v\left(\left|\lambda_{\ell}^{i}\right\rangle\left\langle\lambda_{\ell}^{i}\right|\right)$. We can exploit the linearity condition (2) to show that $R_{i}$ must be independent of $i$. To see this we note that $v\left(E_{1}+E_{2}\right)$ must equal both $\operatorname{Tr}\left(E_{1} R_{1}\right)+\operatorname{Tr}\left(E_{2} R_{2}\right)$ and $\operatorname{Tr}\left[\left(E_{1}+E_{2}\right) R_{12}\right]$ for all possible positive operators $E_{1}$ and $E_{2}$. It follows that our general probability rule is

$$
\begin{equation*}
v\left(E_{i}\right)=\operatorname{Tr}\left(E_{i} R\right) \tag{7}
\end{equation*}
$$

where $R$ is a positive operator that is independent of the $E_{i}$.
If we impose the condition ( P 2 ) then we are led to $\operatorname{Tr}(R)=1$ and so $R$, being a positive unit-trace operator can be understood to be a density operator, but there is no need to race to this result. Rather than imposing the additional property (P2) we can ensure that the probabilities sum to unity by writing

$$
\begin{equation*}
R=\frac{S}{\sum_{j} \operatorname{Tr}\left(E_{j} S\right)} \tag{8}
\end{equation*}
$$

where $S$ is a positive operator and the sum includes all the operators $\left\{E_{j}\right\}$. As $S$ appears both in the denominator and in the numerator of this expression we can, without loss of generality, choose it to have unit trace. This leads us to associate $S$ rather than $R$ with the density operator so that our general probability law is

$$
\begin{equation*}
v\left(E_{i}\right)=\frac{\operatorname{Tr}\left(E_{i} \rho\right)}{\operatorname{Tr}(X \rho)} \tag{9}
\end{equation*}
$$

where $X=\sum_{j} E_{j}$. This is the main result of our paper: if we use only Busch's first and third conditions we arrive a a probability law in which any set of
positive operators (with finite eigenvalues) can provide a set of probabilities and that these probabilities are calculated using (9).

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. Our probability law provides the probabilities for measurement outcomes if we are given the density operator $\rho$. It follows, therefore, that we can understand these as the measurement probabilities given $\rho$ :

$$
\begin{equation*}
v\left(E_{i} \mid \rho\right)=\frac{\operatorname{Tr}\left(E_{i} \rho\right)}{\operatorname{Tr}(X \rho)} \tag{10}
\end{equation*}
$$

Let us consider the effect of a number of possible density operators $\rho_{k}$, prepared with probabilities $p_{k}$ and let $\rho=\sum_{k} p_{k} \rho_{k}$ represent the average or $a$ priori density operator. If we know that the prepared density operator was $\rho_{k}$, then we can use (10) to write

$$
\begin{equation*}
v\left(E_{i} \mid \rho_{k}\right)=\frac{\operatorname{Tr}\left(E_{i} \rho_{k}\right)}{\operatorname{Tr}\left(X \rho_{k}\right)} \tag{11}
\end{equation*}
$$

We should be able to obtain (10) as a sum of these objects, suitably weighted by a probability:

$$
\begin{align*}
v\left(E_{i} \mid \rho\right) & =\sum_{k} v\left(E_{i} \mid \rho_{k}\right) P(k) \\
\Rightarrow \operatorname{Tr}\left(E_{i} \rho\right) & =\sum_{k} \operatorname{Tr}\left(E_{i} \rho_{k}\right) P(k) \frac{\operatorname{Tr}(X \rho)}{\operatorname{Tr}\left(X \rho_{k}\right)} \\
& =\sum_{k} \operatorname{Tr}\left(E_{i} \rho_{k}\right) p_{k} \tag{12}
\end{align*}
$$

For this to hold in general we need only to set

$$
\begin{equation*}
P(k)=\frac{\operatorname{Tr}\left(X \rho_{k}\right)}{\operatorname{Tr}(X \rho)} p_{k} \tag{13}
\end{equation*}
$$

The fact that both the $P(k)$ and the $p_{k}$ are probabilities means that their ratio is a likelihood [9], which we can interpret as the likelihood of $k$ given $X$ :

$$
\begin{equation*}
\ell(k \mid X)=\frac{\operatorname{Tr}\left(X \rho_{k}\right)}{\operatorname{Tr}(X \rho)} \tag{14}
\end{equation*}
$$

In order to adopt this interpretation it is necessary to interpret $P(k)$ as an a posteriori probability based on knowledge gained about the outcome of the measurement. $X$, the sum of operators $E_{j}$ associated with all possible measurement outcomes, is dependent on such knowledge as this can render an outcome impossible that had a non-zero a priori probability of occurring before the measurement. $P(k)$ is then also dependent on this knowledge from Eq.(13). For example, knowledge of which measurement outcome actually occurred makes the a posteriori probability of all other outcomes zero, eliminating their possibility and thereby reducing $X$ to just one term. Another example is where a known post-selection procedure may reject some outcomes from the statistics thereby removing some of the $E_{j}$ allowed by the construction of the measuring device and reducing the number of terms in $X$. We shall express the a posteriori nature of $P(k)$, that is, the probability that $\rho_{k}$ was prepared conditioned on the operator corresponding to the measurement outcome being limited to one of the reduced number of terms in the posterior expression for $X$. This leads us in turn to interpret $v\left(E_{i}\right)$ as

$$
\begin{align*}
v\left(E_{i} \mid \rho, X\right) & =\frac{\operatorname{Tr}\left(E_{i} \rho\right)}{\operatorname{Tr}(X \rho)} \\
& =\sum_{k} \frac{\operatorname{Tr}\left(E_{i} \rho_{k}\right)}{\operatorname{Tr}\left(X \rho_{k}\right)} P(k \mid X) \\
& =\sum_{k} v\left(E_{i} \mid \rho_{k}, X\right) P(k \mid X) \tag{15}
\end{align*}
$$

which is the Bayes rule. There is no need to add a Bayes rule to the usual expression for the quantum probability $\operatorname{Tr}\left(E_{i} \rho\right)$; it is contained already in the general probability law (9).

If there is no post-selection and no knowledge at all about which measurement outcome has occurred, the a posteriori probability $P(k \mid X)$ that $\rho_{k}$ was prepared must be equal to the a priori probability $p_{k}$ for any density operator $\rho_{k}$. In this case we have, from eq. (13)

$$
\begin{equation*}
\operatorname{Tr}\left(X \rho_{k}\right)=\operatorname{Tr}(X \rho) \tag{16}
\end{equation*}
$$

for all $k$. Consider two density operators $\rho_{k}$ with $k=1,2$ related by unitary transformation, $\rho_{2}=U \rho_{1} U^{-1}$. From eq. (16) we then have

$$
\begin{equation*}
\operatorname{Tr}\left(U X \rho_{1} U^{-1}\right)=\operatorname{Tr}\left(X \rho_{1}\right)=\operatorname{Tr}\left(X \rho_{2}\right)=\operatorname{Tr}\left(X U \rho_{1} U^{-1}\right) \tag{17}
\end{equation*}
$$

so $X$ commutes with any $U$ and thus must be proportional to the unit operator, that is $X+K I$, Then $v\left(E_{i}\right)$ in eq. (9) reduces to $\operatorname{Tr}\left(\pi_{i} \rho\right)$ where $\pi_{i}=E_{i} / K$ form a POM and we obtain Born's rule. This leads to condition (P2) which we see is a result of our approach rather than an additional postulate.

It is natural to ask whether there are any applications for our more general probability formula (9). Here we present three such applications. An obvious, but often overlooked, one is to measurement probabilities that are modified 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 [10]. In such cases $X$ will be restricted only to those event operators $E_{i}$ 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 with density operators $\sigma_{i}$ and prior probabilities $q_{i}$ and sends a quantum system prepared in this state to a receiving party, Bob. Bob's task is to determine, as well as possible, the state prepared by Alice. As he knows there is just one outcome, corresponding to $E_{j}$ say, he can simply write $X=E_{j}$ and obtain from (13) the a posteriori, or retrodictive, probability

$$
\begin{equation*}
P\left(i \mid E_{j}\right)=\frac{\operatorname{Tr}\left(E_{j} \sigma_{i}\right.}{\operatorname{Tr}\left(E_{j} \rho\right.} q_{i} \tag{18}
\end{equation*}
$$

We note that because in the general expression (9) $X$ is not assumed to be proportional to the identity operator, there is a symmetry between preparation and measurement which allows Bob to write $E_{i}=q_{i} \sigma_{i}, \rho \propto E_{j}$ and put $X$ equal to the a priori density operator $\sigma=\sum_{i} q_{i} \sigma_{i}$. This gives a result for the probability $v\left(\sigma_{i} \mid E_{j}, \sigma\right)$ that Alice selected the state $\sigma_{i}$, which is the same as (18).

Our final example completes the resolution of a long-standing controversy in retrodictive quantum theory [11]. In retrodictive quantum theory we assign a quantum state on the basis of a later measurement and can use this to ask questions, among other things, about 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 that the prior initial density operator has an unbiased form and is proportional to the identity operator [12, 13]. It is quite straightforward to obtain preparation probabilities, whatever the prior, by applying Bayes' theorem [14]. At a more fundamental level, we can derive a relationship between Bayes' theorem, predictive and retrodictive quantum theory based on an assumed expression for measurement or preparation probabilities in which preparation and measurement operators appear symmetrically [15]. Our general probability rule, however, makes it possible to arrive at the correct expression for retrodictive probabilities without postulating a form for the probabilities. However, it requires a different interpretation to that given earlier. From our general probability law (9) we need only write our retrodictive density operator, based on the measurement outcome $E_{j}$ as

$$
\begin{equation*}
\rho^{r e t r}=\frac{E_{j}}{\operatorname{Tr}\left(E_{j}\right)} \tag{19}
\end{equation*}
$$

to obtain the probabilities for our preparation events $q_{i} \sigma_{i}$ given $\rho^{\text {retr }}$ and $\sigma$ :

$$
\begin{equation*}
v\left(q_{i} \sigma_{i} \mid \rho^{r e t r}, \sigma\right)=\frac{\operatorname{Tr}\left(q_{i} \sigma_{i} \rho^{r e t r}\right.}{\sigma \rho^{r e t r}} \tag{20}
\end{equation*}
$$

which is the required result, obtained previously using either conventional predictive quantum theory supplemented by the Bayes rule or the form of retrodictive quantum theory based on assuming that measurement probabilities are proportional to the trace of the product of a preparation and a measurement operator [15]. Classical probability theory is causally neutral, with prediction and retrodiction linked by Bayes rule. There have been attempts to formulate quantum physics in a causally neutral way [16]. This example shows directly that the general quantum probability rule is also intrinsically causally neutral, with Bayes rule following from it.

We end by summarizing our main conclusions. First and foremost, we have shown that any set of positive operators (with strictly finite eigenvalues) can be used to calculate the probabilities for an event such as the outcome of a measurement. The usually adopted requirement that these operators must sum to the identity is not assumed but follows from our approach for the case where there is no prior information about the measurement outcome. Important examples where we would have such information include quantum communications, retrodictive quantum theory and where there is prior
agreed postselection of measurement results.
This work was supported the Royal Society, the Wolfson Foundation and the UK EPSRC.

## References

1 A. M. Gleason, J. Math. Mech. 6, 885 (1957).
2 A. Peres, Quantum Theory: Concepts and Methods (Kluwer, Dordrecht, 1995).

3 C. J. Isham, Lectures on Quantum Theory (Imperial College Press, London, 1995).

4 P. Busch, Phys. Rev. Lett. 91, 120403 (2003).
5 Gleason's proof applied only to state spaces of dimension greater that two.

6 C. W. Helstrom, Quantum Detection and Estimation Theory (Academic Press, New York, 1976).

7 S. M. Barnett, Quantum Information (Oxford University Press, Oxford, 2009).

8 Busch also adds, to (P3) the conditions that $E+F+\leq 1$, but we drop this supplementary condition along with (P2).

9 G. E. P. Box and G. C. Tiao, Bayesian Inference in Statistical Analysis (Wiley, New York, 1973).

10 P. Grangier, G. Roger and A. Aspect, Europhys. Lett. 1, 173 (1986).
11 S. Watanabe, Rev. Mod. Phys. 27, 179 (1955); Y. Aharonov, P. G. Bergman and J. L. Lebowitz, Phys. Rev. 134, B1410 (1964).

12 F. J. Belinfante, Measurements and time reversal in objective quantum theory (Pergamon Press, Oxford, 1975).

13 T. Amri, J. Laurat and C. Fabre, Phys. Rev. Lett. 106, 020502 (2011).
14 S. M. Barnett, D. T. Pegg and J. Jeffers, J. Mod. Opt. 47, 1779 (2000).

15 D. T. Pegg, S. M. Barnett and J. Jeffers, J. Mod. Opt. 49, 913 (2002).
16 M. S. Leifer and R. W. Spekkens, arXiv:1107.5848 and references therein.

