JEFFREY BUB

VON NEUMANN'S PROJECTION POSTULATE AS A PROBABILITY CONDITIONALIZATION RULE IN QUANTUM MECHANICS

If the transition from classical to quantum mechanics is to be understood as the transition from a Boolean to a non-Boolean possibility structure, several questions arise concerning the representation and interpretation of probabilities, since classical probability theory is essentially a Boolean theory. I want to sketch a representation of classical probability theory as an operator calculus, analogous to the operator calculus of quantum mechanics, and show that the classical conditionalization rule in this calculus is just von Neumann's projection postulate (actually, a corrected form of this postulate first proposed by Lüders).[1] I propose this construction in support of the thesis that von Neumann's projection postulate (more correctly, the Lüders rule) is the appropriate rule for conditionalizing probabilities in the non-Boolean possibility structure of quantum mechanics.

Von Neumann introduces the rule which has become known as the "projection postulate" in Section 3 of Chapter III of his book *Mathematical Foundations of Quantum Mechanics*. In its simplest form, the postulate states that if a measurement of a maximal magnitude A with eigenvalues $a_1, a_2, \ldots$ and corresponding eigenvectors $\alpha_1, \alpha_2, \ldots$ yields the result a_i, then the initial quantum state of the system is transformed to the state α_i. Von Neumann goes on to consider the case of a non-maximal measurement. If the eigenvalue a_i has multiplicity k_i, then the corresponding eigenvectors span a k_i-dimensional subspace K_{a_i}, the range of a projection operator P_{a_i}. Von Neumann argues that after a measurement yielding the result a_i, the system is represented by the statistical operator

$$\frac{P_{a_i}}{\mathrm{Tr}\,(P_{a_i})} = \frac{P_{a_i}}{k_i}.$$

Note that this represents a mixture, not a pure state. In the general case of a magnitude A represented by an operator with a continuous spectrum, he concludes that after a measurement yielding the result $a \in S$, the system is

Journal of Philosophical Logic 6 (1977) 381–390. *All Rights Reserved.*

represented by the unnormalized statistical operator $P_A(S)/P_A(S)$, where $P_A(S)$ is the projection operator in the spectral measure of A corresponding to the range $S(A = \int r\, dP_A(r))$. The operator $P_A(S)$ generates relative probabilities.

Now, quite apart from any objections to a measurement postulate of this sort, it is generally agreed that von Neumann's rule can only be correct for *maximal* measurements. The accepted rule was first proposed by Lüders.[2] The Lüders rule states that a (possibly non-maximal) measurement of a magnitude A yielding the result a_i leads to the following transition in the statistical operator W of the system:

$$W \rightarrow W' = \frac{P_{a_i} W P_{a_i}}{\mathrm{Tr}\,(P_{a_i} W P_{a_i})} \qquad \text{(Lüders)}^3$$

and *not*

$$W \rightarrow W' = \frac{P_{a_i}}{\mathrm{Tr}\,(P_{a_i})} \qquad \text{(von Neumann)}$$

What is the difference between these two rules? The two rules agree in only two cases: (i) if $W = I$(unnormalized) where I is the unit operator, and (ii) for maximal measurements, i.e., when each P_{a_i} is the projection operator onto a (different) 1-dimensional subspace spanned by the vector α_i.

Case (i) is immediately obvious:

$$W \rightarrow W' = \frac{P_{a_i}}{\mathrm{Tr}\,(P_{a_i})}.$$

For case (ii), the von Neumann rule yields:

$$W \rightarrow W' = P_{\alpha_i}$$

for the transition corresponding to the result a_i. The Lüders rule yields this transition too:

$$W \rightarrow W' = \frac{P_{\alpha_i} W P_{\alpha_i}}{\mathrm{Tr}\,(P_{\alpha_i} W P_{\alpha_i})} = P_{\alpha_i}.$$

(Note that $P_{\alpha_i} W P_{\alpha_i} = (\alpha_i, W\alpha_i) P_{\alpha_i} = \mathrm{Tr}\,(W P_{\alpha_i}) P_{\alpha_i}$.)

To bring out the difference between the Lüders rule and the von Neumann Neumann rule, consider an initial pure statistical operator $W = P_\psi$ and a non-maximal measurement, i.e., where the projection operators are not in general 1-dimensional.

The Lüders rule yields

$$W = P_\psi \to W' = \frac{P_{a_i} P_\psi P_{a_i}}{\mathrm{Tr}\,(P_{a_i} P_\psi P_{a_i})}.$$

Let $P_{a_i}\psi = \theta_i$, then for any vector ϕ:

$$\begin{aligned} P_{a_i} P_\psi P_{a_i}\phi &= P_{a_i} P_\psi (P_{a_i}\phi) \\ &= P_{a_i}(\psi, P_{a_i}\phi)\psi \\ &= (P_{a_i}\psi, \phi)\theta_i \\ &= (\theta_i, \phi)\theta_i \\ &= P_{\theta_i}\phi \end{aligned}$$

i.e. $$P_{a_i} P_\psi P_{a_i} = P_{\theta_i}.$$

Now, $$\mathrm{Tr}\,(P_{a_i} P_\psi P_{a_i}) = \|P_{a_i}\psi\|^2 = \|\theta_i\|^2,$$

and so $$W' = P_{\theta_i}/\|\theta_i\|^2 = P_{\theta'_i},$$

where $$\theta'_i = \frac{\theta_i}{\|\theta_i\|}.$$

The von Neumann rule yields

$$W = P_\psi \to W' = \frac{P_{a_i}}{\mathrm{Tr}\,(P_{a_i})} \quad \text{or } W' = P_{a_i} \text{ (unnormalized) if } \mathrm{Tr}(P_{a_i}) = \infty$$

According to the von Neumann rule, W' is a *mixture which does not depend on the initial quantum state* ψ. According to the Lüders rule, W' is a *pure state* θ'_i, *which does depend on the initial state* ψ. In fact, θ'_i is the normalized projection of ψ onto the subspace which is the range of P_{a_i}, i.e., $\theta'_i = \dfrac{P_{a_i}\psi}{\|P_{a_i}\psi\|}$.

To show that the Lüders rule, and not the von Neumann rule, is the appropriate rule for conditionalizing probabilities in the non-Boolean possibility structure of quantum mechanics, consider for simplicity a countable classical probability space $(X, \mathcal{F}, \mu)$. I shall label the atomic events or elementary possibilities by $x_1, x_2, \ldots$. These are associated with singleton subsets $X_1, X_2, \ldots$, or indicator functions (characteristic functions) $I_1, I_2, \ldots$. I shall label other, possibly non-atomic, events by $a, b, \ldots$. Thus, the set $a_1, a_2, \ldots$ might denote a set of non-atomic mutually exclusive and collectively exhaustive events ($\sum_i I_{a_i} = I; I_{a_i} I_{a_j} = 0, i \neq j$).

Now, for any probability measure μ, it is possible to introduce a "statistical operator" $W = \Sigma_i w_i I_i$, where $\Sigma_i w_i = 1$, $w_i \geqslant 0$, for all i, in terms of which the probability of an event a may be represented as:

$$p_\mu(a) = \mu(X_a) = \sum_j \left(\sum w_i I_i(x_j)\right) I_a(x_j).$$

I shall write $p_W(a)$ for $p_\mu(a)$, where W corresponds to μ, i.e.

$$p_W(a) = \sum_j W(x_j) I_a(x_j)$$

To simplify notation, I shall abbreviate this expression as

$$p_W(a) = \sum W I_a$$

where a summation sign without an index is understood as summing over all the atomic events x_j. This convention will be used below.

In terms of the statistical operator, the conditional probability (relative to an initial measure μ associated with the statistical operator W) of an event b given an event a_i, may be represented as:

$$p_W(b\,|\,a_i) = \frac{\Sigma\, W I_{a_i} I_b}{\Sigma\, W I_{a_i}}$$

(i.e.,

$$\left. p_W(b\,|\,a_i) = \frac{\Sigma_j\, W(x_j) I_i(x_j) I_b(x_j)}{\Sigma_j\, W(x_j) I_i(x_j)} \right).$$

To see this, simply notice that

$$p_\mu(b\,|\,a_i) = \frac{\mu(X_{a_i} \cap X_b)}{\mu(X_{a_i})} = \frac{\Sigma\, W I_{a_i} I_b}{\Sigma\, W I_{a_i}}.$$

Thus, the transition

$$\mu \to \mu'$$

on conditionalization with respect to a_i (where μ' is defined by $\mu'(X_e) = \mu(X_{a_i} \cap X_e)/\mu(X_{a_i}) = p_\mu(e\,|\,a_i)$ for any event e) may be represented as the transition

$$W \to W' = \frac{WI_{a_i}}{\Sigma\, WI_{a_i}}$$

so that

$$p_W(b\,|a_i) = \Sigma\, W'I_b.$$

The statistical operator construction allows the replacement of the measure function μ, which is a set function whose domain is the field of measurable subsets of X, by a corresponding random variable W, a point function whose domain is X. If we regard the probability space as associated with a physical system with magnitudes A, B, etc., whose possible values $a_1, a_2, \ldots; b_1, b_2, \ldots$; etc. correspond to the possible events represented by the field $\mathfrak{F}$, then the statistics of this system is now represented by a physical magnitude W belonging to the algebra of magnitudes of the system. In fact, W is a linear combination of atomic magnitudes of the system. The advantage of this construction is that it provides a purely algebraic way of representing the statistics of a system, which is appropriate whether or not a representation of the algebra of magnitudes as real-valued functions on a space is possible. I want to suggest that we take W as representative of the statistics in a primary sense – the measure function μ exists only if the algebra of magnitudes is commutative. In this special case (a classical probability space), the subalgebra of idempotent magnitudes forms a Boolean algebra, which has a representation as a field of subsets of a set, by Stone's theorem. The measure function defined as a set function on this field is essentially the "Stone representative" of the statistical operator W, which is the element in the algebra of magnitudes incorporating the statistics. Bearing in mind the possibility of non-commutative algebras of magnitudes as in quantum mechanics (i.e., non-Boolean possibility structures), it seems appropriate to represent the transition corresponding to conditionalization with respect to an event a_i by the symmetrical expression:

$$W \to W' = \frac{I_{a_i}WI_{a_i}}{\Sigma\, I_{a_i}WI_{a_i}}.$$

Now, *this is just the Lüders version of von Neumann's projection* postulate. In quantum mechanics, the statistics of a system is represented by a statistical operator W which may be represented as:

$$W = \sum_i w_i P_{\alpha_i}$$

where the P_{α_i} are projection operators onto atomic events (i.e., projection operators onto 1-dimensional subspaces spanned by the vectors α_i). Thus, W is a linear combination of atomic idempotent magnitudes of the system. In terms of this operator, the probability of an event b is represented as:

$$p_W(b) = \mathrm{Tr}\,(WP_b).$$

Notice that the trace of an operator O is just the sum of the eigenvalues of O, i.e., the sum of the possible values of O at each atom in the maximal Boolean subalgebra defined by O. Thus, the operation Tr in the non-commutative algebra of magnitudes of a quantum mechanical system is completely analogous to the operation Σ in the commutative algebra of magnitudes of a classical mechanical system.

According to the Lüders rule, the conditional probability (relative to an initial measure associated with the statistical operator W) of an event b given an event c_i is:

$$p_W(b\,|\,c_i) = \mathrm{Tr}\,(W'P_b)$$

where

$$W' = \frac{P_{c_i}WP_{c_i}}{\mathrm{Tr}\,(P_{c_i}WP_{c_i})}.$$

If we assume an *a priori* probability assignment given by the unnormalized statistical operator $W = I$, representing an equiprobable initial distribution over every complete set of orthogonal atomic properties (associated with the possible values of a maximal magnitude), then conditionallization with respect to an atomic property c_i yields the transition

$$W \to W' = P_{c_i}$$

where P_{c_i} is the projection operator onto the 1-dimensional subspace spanned by the eigenvector γ_i, say, corresponding to c_i. This means that the probability of a property b conditional on c_i (where b may be incompatible with c_i) is to be computed according to the rule:

$$p_W(b\,|\,c_i) = \mathrm{Tr}\,(P_{c_i}P_b).$$

If b is atomic, corresponding to the vector β, we have

$$p_W(b|c_i) = |(\beta, \gamma_i)|^2.$$

Thus, the probability assigned by the "state vector" γ_i (representing the association of the property c_i with the system) to an incompatible property b, according to the quantum mechanical rule, may be interpreted as the conditional probability of the property b given the property c_i relative to an initial probability distribution with is equiprobable with respect to every complete set of atomic properties of the system.

A number of problems of interpretation are resolved if we interpret the quantum mechanical specification of a system by its state vector as a statistical specification in the above sense, in which the statistical operator P_ψ determined by the state vector ψ is understood as the algebraic counterpart of the classical measure function μ, which does not exist in this case since the possibility structure is non-Boolean. (Of course, the state vector cannot be interpreted in this way if the possible values of the maximal quantum mechanical magnitudes are represented as generating a classical probability space – whatever initial probability distribution we choose.)

For example, the 2-slit experiment can be analyzed as a problem in conditional probabilities on a non-Boolean possibility structure. It can be shown that the von Neumann rule gives the wrong result (no interference) while the Lüders rule gives the correct result. We have a screen with two slits, A and B, and a second detecting screen or photographic plate. A photon in a pure quantum state represented by a plane wave moves towards the slits. Each slit can be regarded as localizing the photon to a region, Δ_A or Δ_B, in the plane of the slit screen. In other words, there is a magnitude M, representing position in the slit screen plane, and the passage of a particle through a slit is a measurement of the magnitude M, in the sense that a range range, Δ_A or Δ_B, is assigned to M for the photon at the time of passage. We are interested in the probability that the photon will arrive at a certain region on the detecting screen, conditional on localization to a certain range of values of M (Δ_A, Δ_B, or $\Delta_A \cup \Delta_B$). Localization to a region Δ on the detecting screen is a measurement of a magnitude N, representing position in the detecting screen plane. N may be taken as the same magnitude M, if the regions are the same size as the slits, or at least as compatible with M otherwise.

It is easy to show that the conditionalized statistical operator for the photon, immediately after the photon has passed through the slit system with both slits open, is:

$$W_{AB} = P_M(\Delta_A) + P_M(\Delta_B) \quad \text{(von Neumann, unnormalized)}$$

$$W_{AB} = P_\theta \quad \text{(Lüders)}$$

where $P_M(\Delta_A), P_M(\Delta_B)$ are the projection operators in the spectral measure of M corresponding to the ranges Δ_A, Δ_B, respectively, and $\theta = (\psi_A + \psi_B)/\sqrt{2}$, where ψ_A is the normalized projection of the initial state ψ (plane wave) onto the subspace which is the range of the projection operator $P_M(\Delta_A)$ (and similarly for ψ_B).

Notice that according to the von Neumann rule W_{AB} represents a mixture which does not in any way depend on the initial quantum state of the photon. But the initial state is *required* to be a plane wave (and not, say, a mixture of plane waves, as might be obtained by placing a candle to the left of the slit screen) in order to obtain interference effects. It follows that the probability of the photon arriving at region Δ on the detecting screen with both slits open is simply one half the sum of the probabilities with either slit A or slit B open, irrespective of the distance between the slit screen and the detecting screen. This is what one would expect on a classical analysis, and is contradicted by the interference pattern.

According to the Lüders rule, W_{AB} is a pure statistical operator, and the probability of the photon arriving at region Δ on the detecting screen with both slits open is equal to one half the sum of the probabilities with either slit A or slit B open only when the slit screen and the detecting screen are zero distance apart. For a non-zero distance between the screens, the probability assigned by the time-evolved statistical operator $W_{AB} = P_\theta$ differs from the sum by a non-zero "interference term".

This analysis of the 2-slit experiment makes clear the role played by (i) the initial quantum state, and (ii) the non-zero distance between the slit screen and the detecting screen. The explanation of the interference effect depends on the difference between the Lüders conditionalization rule and the von Neumann rule. Von Neumann's rule:

$$W \to W' = \frac{P_{a_i}}{\mathrm{Tr}\,(P_{a_i})} \quad \text{(assuming } \mathrm{Tr}(P_{a_i}) < \infty)$$

is the analogue of the classical rule:

$$W \to W' = \frac{I_{a_i}}{\Sigma\,(I_{a_i})}$$

representing a conditionalization and *randomization* of the initial measure within the subset X_{a_i} – i.e., the initial measure is not merely "renormalized" to include the new information (that the value of A is a_i), but replaced by a *uniform* measure over the set X_{a_i} (so that information contained in the initial measure *concerning the relative probabilities of properties represented by subsets in* X_{a_i} is lost). Thus, the "paradox" involved in the 2-slit experiment is resolved by showing precisely how the assumption of a non-Boolean possibility structure explains the existence of the "anomalous" interference effect.[4]

On the usual interpretation, the projection postulate is a rule representing the effect of the necessarily finite and uncontrollable disturbance of a system involved in any quantum mechanical measurement process. My point is that the projection postulate in its corrected Lüders version is properly understood as *mere conditionalization on a non-Boolean possibility structure*, since it is the analogue of mere conditionalization on new information in the Boolean case. The effect of a measurement disturbance involved in obtaining this information would be represented as an *additional* change in the statistical operator, over and above the change defined by the Lüders rule. Such a measurement disturbance may be more or less violent. The von Neumann rule corresponds to the most violent disturbance possible, in which *all* initial information concerning the system is lost, and only information represented by the measurement result is retained.[5]

A similar analysis of the Einstein–Podolsky–Rosen experiment as a problem in conditional probabilities shows how the peculiar quantum mechanical correlations between conditional probabilities involving the separated systems S and S' arise on the non-Boolean possibility structure. In this case, again, conditionalization with respect to non-atomic properties is involved (an S-magnitude or S'-magnitude is non-maximal in the Hilbert space of the composite system $S + S'$) so that the difference between the Lüders and von Neumann versions of the projection postulate comes into play.

In general, then, insofar as problems of interpretation in quantum mechanics have their source in probability relations which are anomalous classically, these problems are resolved by recognizing the projection postulate in its corrected Lüders version as the appropriate conditionalization rule on the non-Boolean possibility structure of a quantum mechanical system. If the logical interpretation of quantum mechanics can make sense of truth, it can make sense of probability.

University of Western Ontario

NOTES

[1] A non-commutative extension of the classical notion of conditional probability (more generally, conditional expectation) has been extensively investigated by Umegaki (H. Umegaki, 'Conditional Expectation in an Operator Algebra', I. *Tohoku Math. J.* 6, 177–181 (1954); II, 8, 86–100 (1956); III. *Kodai Math. Semi. Rep.* 11, 51–64 (1959); IV. *Kodai Math. Semi. Rep.* 14, 59–85 (1962)). Umegaki's theory has recently been extended to magnitudes with continuous spectra by Davies and Lewis (E. B. Davies and J. T. Lewis, *Commun. Math. Phys.* 17, 239–260 (1970)). Using the Umegaki theory, Nakamura and Umegaki have shown that von Neumann's projection postulate is just the conditionalization of the statistical operator relative to an event in the non-Boolean possibility structure. (M. Nakamura and H. Umegaki, 'On von Neumann's Theory of Measurements in Quantum Statistics', *Math. Jap.* 7, 151–157 (1961–62)). Their demonstration considers only maximal (i.e., non-degenerate) magnitudes with discrete spectra, in which case the Lüders rule coincides with von Neumann's rule. A discussion of the Lüders rule *vis a vis* von Neumann's rule follows.

[2] G. Lüders, *Ann. d. Physik* 8, 322 (1951). The Lüders rule is discussed at some length by Furry in W. H. Furry, 'Some Aspects of the Quantum Theory of Measurement', *Lectures in Theoretical Physics Volume VIII A, Statistical Physics and Solid State Physics,* University of Colorado Press, Boulder, 1966.

[3] By the properties of the trace operation,

$$\mathrm{Tr}\,(P_{a_i}WP_{a_i}) = \mathrm{Tr}\,(WP_{a_i}) = \mathrm{Tr}\,(P_{a_i}W).$$

I shall continue to write such expressions in symmetrical form below.

[4] Compare this analysis with Putnam's discussion in H. Putnam, 'Is Logic Empirical?', *Boston Studies in the Philosophy of Science*, R. Cohen and M. Wartofsky (eds.), Reidel, 1969. Putnam's solution to the problem posed by the phenomenon of interference is to block the application of the distributive law in transforming the conditional probability on passage of the photon through both slits, to a sum of conditional probabilities for each of the slits separately. This solution is spurious, however, because the usual classical notion of conditional probability is inapplicable if the possibility structure is non-Boolean. Notice that the initial quantum state plays no role in Putnam's analysis, and there is no explicit recognition of the significance of the distance between the slit screen and the detecting screen (although a non-zero distance is implicitly required for the non-distributivity of the events considered).

[5] At first sight it might seem that the application of the von Neumann rule is appropriate in the case of the 2-slit experiment when some physical device is incorporated into the experimental arrangement for detecting the passage of an individual photon through either slit A or slit B exclusively, when both slits are open. Now, while it might indeed be the case that certain devices of this sort introduce a disturbance represented by the von Neumann rule, there is no *theoretical* reason I can see why such devices should not operate with minimal disturbance. In this case, both the von Neumann rule and the Lüders rule yield the same result: no interference. On the von Neumann rule, the statistical operator of the photon immediately after passing through the slit system with the detection device is the mixture:

$$W_{AB} = P_M(\Delta_A) + P_M(\Delta_B) \quad \text{(unnormalized)}$$

On the Lüders rule, the statistical operator is the mixture:

$$W_{AB} = \tfrac{1}{2}P_{\psi_A} + \tfrac{1}{2}P_{\psi_B}.$$

Both mixtures yield no interference at the detecting screen, for any distance between the slit screen and the detecting screen.