ANNALS OF PHYSICS **192**, 368–382 (1989)

How Probability Arises in Quantum Mechanics*

EDWARD FARHI AND JEFFREY GOLDSTONE

Center for Theoretical Physics,
Laboratory for Nuclear Science and Department of Physics,
Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139

AND

SAM GUTMANN

Department of Mathematics, Northeastern University,
Boston, Massachusetts 02115

Received February 2, 1989

DEDICATED TO HERMAN FESHBACH ON HIS 70TH BIRTHDAY

A version of the postulates of quantum mechanics is presented in which no reference is made to probability. Instead we rely on a weaker postulate referring to eigenvalues and eigenstates. The modulus squared of the inner product of two state vectors is shown to be an eigenvalue of the operator representing a *frequency* measurement on the system of an infinite number of copies of the original system. The argument makes essential use of the Strong Law of Large Numbers. © 1989 Academic Press, Inc.

I. INTRODUCTION

We all believe that quantum mechanics, as opposed to classical mechanics, gives a probabilistic description of nature [1]. The probabilistic interpretation of measurement is contained in one of the standard postulates of quantum mechanics. In this paper we will strengthen the argument that the postulates of quantum mechanics can be weakened so that they make no reference to probability. The probabilistic interpretation will arise naturally from the weakened postulates without having been put in beforehand.

To begin we list the standard postulates of quantum mechanics. They are:

PI: The states of a quantum system, S, are described by vectors $|\psi\rangle$ which are elements of a Hilbert space, V, that describes S.

* This work is supported in part by funds provided by the U.S. Department of Energy (DOE) under Contract DE-AC02-76ER03069.

0003-4916/89 $7.50

PII: The states evolve in time according to

$$H\,|\psi\rangle = i\hbar\frac{\partial}{\partial t}|\psi\rangle,$$

where H is a Hermitian operator which specifies the dynamics of the system S.

PIII: Every observable, $\mathcal{O}$, is associated with a Hermitian operator θ. The only possible outcome of a measurement of $\mathcal{O}$ is an eigenvalue θ_i of θ.

$$\theta\,|\theta_i\rangle = \theta_i\,|\theta_i\rangle; \qquad \langle\theta_i|\theta_j\rangle = \delta_{ij}.$$

PIV: If the state of the system is described by the normalized vector $|\psi\rangle$, then a measurement of $\mathcal{O}$ will yield the value θ_i with probability

$$p_i = |\langle\theta_i|\psi\rangle|^2.$$

In order for successive measurements of $\mathcal{O}$ to yield the same value θ_i it is necessary to have the projection postulate:

PV: Immediately after a measurement which yields the value θ_i the state of the system is described by $|\theta_i\rangle$.

It is well known that the type of time evolution implied by PV is incompatible with the unitary time evolution implied by PII. We will have more to say about this at the end of our paper. For now we will advocate the point of view that PIV can be replaced by the weaker postulate:

PIV′: If a quantum system is described by the state $|\theta_i\rangle$ then a measurement of $\mathcal{O}$ will yield the value θ_i.

Clearly PIV′ is a special case of PIV but it is *not* a statement about probabilities. The replacement of PIV by PIV′ eliminates the immediate need for PV since the state is $|\theta_i\rangle$ before and after the measurement.

In order to extract the concept of probability from PIV′ (we also need PI and PIII) we consider a large system, **S**, which consists of identical copies of the system S. To begin, we consider a finite number of copies, N. Consider the observable $f^{(N)}(\theta_i)$ which is the fraction of the N systems which are measured to have the value θ_i when θ is measured. There is an associated frequency operator $F^{(N)}(\theta_i)$ which acts in the Hilbert space of **S**, $V^{(N)} = V \otimes V \otimes \cdots \otimes V$. It can be shown that the state $|\psi\rangle^{(N)} = |\psi\rangle \otimes |\psi\rangle \otimes \cdots \otimes |\psi\rangle$ obeys

$$\lim_{N\to\infty} \| F^{(N)}(\theta_i)|\psi\rangle^{(N)} - p_i\,|\psi\rangle^{(N)}\| = 0,$$

where $p_i = |\langle\theta_i|\psi\rangle|^2$. It is tempting to take this to mean that the state consisting of an infinite number of copies of $|\psi\rangle$ is an exact eigenstate of the frequency operator with eigenvalue p_i. Then using PIV′ we would conclude that a measurement of the proportion of systems in which we observe θ_i would yield p_i and the probabilistic interpretation of the inner product $\langle \theta_i | \psi \rangle$ would be derived.

The trouble with this reasoning is that a statement about the limit as N goes to infinity is really a statement about finite N. It says that given any $\varepsilon > 0$, there is a finite N such that $\| F^{(N)}(\theta_i) | \psi \rangle^{(N)} - p_i | \psi \rangle^{(N)} \| < \varepsilon$, which is not an exact eigenstate equation. Therefore, postulate PIV′ does not apply and no conclusion can be reached. To remedy this we will start directly with the large system **S** consisting of an infinite number of copies of S. To describe this fully we need a non-separable Hilbert space. However, we can sharpen the discussion by restricting our attention to one separable component of the Hilbert space which contains the vector $| \psi \rangle^{\infty}$ which represents an infinite number of copies of $| \psi \rangle$. Without considering a limit of frequency operators $F^{(N)}$ we will define a frequency operator, $F(\theta_i)$, and we will prove

$$F(\theta_i) | \psi \rangle^{\infty} = p_i | \psi \rangle^{\infty},$$

where again $p_i = | \langle \theta_i | \psi \rangle |^2$. We can then apply PIV′ to this exact eigenvector equation and the probabilistic interpretation follows.

The above discussion has direct parallels with the classical probability Weak Law of Large Numbers and Strong Law of Large Numbers. The Weak Law of Large Numbers says that if the probability of a coin flip turning up heads is p, then you can always find an N large enough—N being the number of trials—that the probability that the proportion of heads differs from p by more than a fixed amount δ is always less than ε, for any ε and δ. In other words if you flip an ordinary coin the probability that you do not get between 49 and 51% heads can be made as small as you like if you flip enough coins. By making this probability very small, the Weak Law can be construed to have empirical content. However, since the term probability still appears in the conclusion there are conceptual problems about what it really means.

The probabilist's way to deal with this is to consider infinite sequences of coin flips. There are an uncountable number of such sequences and a probability measure can be assigned to this space of sequences. The probability measure has the property that the probability of the set of all sequences which begin with a definite string of k heads and l tails is $p^k(1-p)^l$. This property determines the measure completely. The Strong Law of Large Numbers says that the set of all sequences whose proportion of heads is p has probability measure one. In other words those sequences which have no definite proportion of heads or whose proportion differs from p are a set of measure zero. The Strong Law is a statement about infinite sequences, not limits, and it implies the Weak Law.

The statement that $| \psi \rangle^{\infty}$ is an eigenstate of the frequency operator, $F(\theta_i)$, with eigenvalue p_i is a quantum version of the Strong Law of Large Numbers. It is a precise statement but because it deals with an infinite number of copies we have lost direct contact with experimental reality.

The idea that probability can be viewed as a consequence of quantum mechanics without postulates PIV and PV was discussed by Everett [2]. The frequency operator was introduced by Graham [3] and Hartle [4]. Hartle also gives a Weak Law version of our results.

What is novel in our approach is that our frequency operator is *not* defined as the limit of finite frequency operators. Rather its properties, and hence our probability interpretation, depend crucially on a probability measure. This probability measure in turn is forced on us by the necessity to describe physics in a basis independent manner.

II. Classical Probability—The Weak Law of Large Numbers

Consider a sequence of independent trials where the outcome of the rth trial is j_r, which takes the values 0, 1 with probabilities p_0, p_1. The probability of the finite sequence $j_1, j_2, ..., j_N$ is $\prod_{r=1}^{N} p_{j_r}$. Define $f^{(N)}(j_1, j_2, ..., j_N) = (1/N) \sum_{r=1}^{N} j_r$ as the frequency of the outcome $j = 1$. Then averaging over all sequences weighted by their probability gives

$$\langle (f^{(N)} - p_1)^2 \rangle = \frac{1}{N} p_0 p_1 \to 0 \qquad \text{as} \quad N \to \infty.$$

For any random variable, x, and any probability distribution, it is easy to show that $\langle x^2 \rangle \geqslant \delta^2 \operatorname{Prob}[|x| > \delta]$. It follows that for any $\delta > 0$,

$$\operatorname{Prob}[|f^{(N)} - p_1| > \delta] \to 0 \qquad \text{as} \quad N \to \infty, \tag{WI}$$

which is known as The Weak Law of Large Numbers [5]. In order to assert WI we consider only *finite* sequences of independent trials. We can also state (WI) as follows: Given any $\varepsilon > 0$, $\delta > 0$, there exists $N_0(\varepsilon, \delta)$ such that

$$\begin{aligned} &\operatorname{Prob}[|f^{(N_0+1)} - p_1| > \delta] < \varepsilon, && \text{and} \\ &\operatorname{Prob}[|f^{(N_0+2)} - p_1| > \delta] < \varepsilon, && \text{and} \cdots \\ &\operatorname{Prob}[|f^{(N_0+M)} - p_1| > \delta] < \varepsilon && \text{for any } M > 0. \end{aligned} \tag{WII}$$

III. Quantum Mechanics—The Weak Frequency Law

Consider a simple quantum system S described by a two-dimensional Hilbert space V. (The case when V is n-dimensional is described in Appendix C.) Let A and B each be operators with eigenvalues 0, 1 and eigenvectors $|a, i\rangle$, $|b, j\rangle$ so that $A|a, i\rangle = i|a, i\rangle$, $B|b, j\rangle = j|b, j\rangle$; $i, j = 0, 1$. Now $|a, i\rangle$ and $|b, j\rangle$ each form an orthonormal basis for V. Let $p_j = |\langle b, j | a, 0\rangle|^2$, so $p_0 + p_1 = 1$. In the standard interpretation p_j is the probability of finding the eigenvalue j when we measure B in the eigenstate $|a, 0\rangle$ of A, but we will not appeal to this interpretation.

Consider the Hilbert space

$$V^{(N)} = V \otimes V \otimes \cdots \otimes V,$$

which describes N copies of the system S. The operators $A_{(r)}$, $B_{(r)}$ are A and B operating on the rth factor, i.e., $A_{(r)} = I \otimes \cdots \otimes I \otimes A \otimes I \otimes \cdots \otimes I$. The simultaneous eigenstates of $A_{(1)} \cdots A_{(N)}$ with eigenvalues $i_1 \cdots i_N$ are

$$|a; i_1 \cdots i_n\rangle = |a, i_1\rangle \otimes |a, i_2\rangle \otimes \cdots \otimes |a, i_N\rangle$$

and form a complete set of 2^N orthonormal vectors, a basis for $V^{(N)}$. Similarly define $|b; j_1 \cdots j_N\rangle$. Define the operator $F^{(N)}$ by $F^{(N)} = (1/N) \sum_{r=1}^{N} B_{(r)}$, i.e.,

$$F^{(N)} |b; j_1 \cdots j_N\rangle = \left(\frac{1}{N} \sum_{r=1}^{N} j_r\right) |b; j_1 \cdots j_N\rangle.$$

$F^{(N)}$ is the operator for the observable which is the frequency of the outcome $j = 1$. Now

$$\langle a; 0 \cdots 0 | (F^{(N)} - p_1)^2 \, |a; 0 \cdots 0\rangle$$
$$= \sum_{j_r = 0,\, 1} \prod_{r=1}^{N} p_{j_r} (f^{(N)}(j_1 \cdots j_N) - p_1)^2 = \frac{p_0 p_1}{N},$$

exactly as in Section II. Thus the quantum Weak Frequency Law is the statement

$$\|(F^{(N)} - p_1) |a; 0 \cdots 0\rangle\| \to 0 \qquad \text{as} \quad N \to \infty;$$

i.e., for large enough N, $|a; 0 \cdots 0\rangle$ is *almost* an eigenstate of $F^{(N)}$. This, however, is not good enough to replace postulate PIV. For that, we need a suitable translation of the classical Strong Law of Large Numbers.

IV. Classical Probability—The Strong Law of Large Numbers

Suppose we are willing to assign probabilities to events consisting of sets of infinite sequences of outcomes $\{j_1, j_2, \ldots\}$, $j_r = 0, 1$. We clearly want the probability of the set of all sequences beginning $j_1, j_2, \ldots, j_N$ to be $\prod_{r=1}^{N} p_{j_r}$. Modern probability theory tell us that there is a unique probability measure with this property. The Strong Law of Large Numbers [5], which we state without proof, asserts that

$$\text{Prob}\, [\lim_{N \to \infty} f^{(N)}(j_1, j_2, \ldots, j_N) = p_1] = 1. \tag{SI}$$

This means that the set of all sequences whose ultimate proportion of 1's is not p_1

has probability zero. In contrast to the weak laws, (SI) is a statement about *infinite* sequences of trials.

In the course of the proof of (SI), one does demonstrate a statement about probabilities of finite sequences of the same nature as (WII), but manifestly stronger. Given any $\varepsilon > 0$, $\delta > 0$, there exists $N_0(\varepsilon, \delta)$ such that for any $M > 0$

$$\text{Prob}[|f^{(N_0+1)} - p_1| > \delta \quad \text{or} \quad |f^{(N_0+2)} - p_1| > \delta, ..., \quad \text{or} \quad |f^{(N_0+M)} - p_1| > \delta] < \varepsilon. \quad \text{(SII)}$$

However, it is (SI) we need in order to address the conceptual problem of probability and we need a quantum equivalent of (SI) to eliminate reference to probability in the postulates of quantum mechanics. (For a translation of (SII) into operator language see Ref. [6].)

V. The Quantum System of an Infinite Number of Copies

We wish to construct a Hilbert space which describes an infinite number of copies of S. The full Hilbert space is non-separable, i.e., it has an uncountable orthonormal basis. However, we will focus our attention on a single component of the big Hilbert space which contains the vector $|a; \{0, 0, ...\}\rangle$. We define this component, $V_c^{(\infty)}$, which itself is a Hilbert space, as follows. An orthonormal basis is $|a; \{i\}\rangle$, where $\{i\} = \{i_1, i_2, ...\}$ is an infinite sequence of zeros and ones with a *finite* number of ones. (In our notation $\{\ \}$ is reserved for infinite sequences.) Any state can be expanded.

$$|\Psi\rangle = \sum_{\{i\}} \langle a; \{i\} | \Psi\rangle |a; \{i\}\rangle$$

with

$$\sum_{\{i\}} |\langle a; \{i\} | \Psi\rangle|^2 < \infty.$$

Since the set of sequences $\{i\}$ is countable, this is just the Hilbert space l_2 of square summable sequences. The operator $A_{(r)}$ acts on this space by $A_{(r)}|a; \{i\}\rangle = i_r\, |a; \{i\}\rangle$. The matrix elements of the operator $B_{(r)}$ are given by

$$\langle a; \{i'\} | B_{(r)} | a; \{i\}\rangle = \langle a, i'_r | B | a, i_r\rangle \prod_{s \neq r} \delta_{i'_s i_s}.$$

We can describe observations on any *finite* number of copies using the space $V_c^{(\infty)}$. Clearly the operators $B_{(1)}, B_{(2)}, ...$ all commute. A simultaneous eigenstate $|\Psi\rangle$ of $B_{(1)} \cdots B_{(N)}$ with eigenvalues $j_1 \cdots j_N$ must have

$$\langle a; \{i\} | \Psi\rangle = \prod_{r=1}^{N} \langle a, i_r | b, j_r\rangle\, \hat{\psi}_N(i_{N+1}, ...), \qquad (*)$$

where $\hat{\psi}_N$ is a function which depends only on $i_{N+1}, i_{N+2}, \ldots$. Can we write down a simultaneous eigenstate of *all* $B_{(r)}$ with eigenvalues $j_1, j_2, \ldots$, i.e., a state $|\Psi\rangle$ such that $B_{(r)}|\Psi\rangle = j_r|\Psi\rangle$, for an *infinite* sequence $\{j\}$ of zeros and ones? Such a $|\Psi\rangle$ must satisfy ($*$) for all N. For any sequence $\{i\}$ there exists an r_1 such that $i_r = 0$ for $r > r_1$ since there are only a finite number of 1's in $\{i\}$. Then

$$\frac{\langle a; \{i\} | \Psi\rangle}{\langle a; \{0\} | \Psi\rangle} = \prod_{r=1}^{r_1} \frac{\langle a, i_r | b, j_r\rangle}{\langle a, 0 | b, j_r\rangle}$$

and so

$$\langle a; \{i\} | \Psi\rangle = \langle a; \{0\} | \Psi\rangle \prod_{r=1}^{\infty} \frac{\langle a, i_r | b, j_r\rangle}{\langle a, 0 | b, j_r\rangle}.$$

The last product is well-defined since each factor is 1 for r large enough. Now we calculate

$$\begin{aligned}\sum_{\{i\}} |\langle a; \{i\} | \Psi\rangle|^2 &= \lim_{N\to\infty} \sum_{i_1, i_2, \ldots, i_N} |\langle a; \{i_1 \cdots i_N, 0, 0, \ldots\} | \Psi\rangle|^2 \\ &= \lim_{N\to\infty} |\langle a; \{0\} | \Psi\rangle|^2 \prod_{r=1}^{N} \frac{\sum_{i_r=0,1} |\langle a, i_r | b, j_r\rangle|^2}{|\langle a, 0 | b, j_r\rangle|^2} \\ &= \frac{|\langle a; \{0\} | \Psi\rangle|^2}{\prod_{r=1}^{\infty} |\langle a, 0 | b, j_r\rangle|^2}.\end{aligned}$$

Of course, the product in the denominator is zero, since each factor is p_0 or p_1 and $p_0 < 1$, $p_1 < 1$. That is, there is no normalizable state $|\Psi\rangle$. This is no surprise; we are attempting to construct a basis for $V_c^{(\infty)}$ corresponding to the *uncountable* set of *all* sequences $\{j\}$ of zeros and ones. At this point it is clear that we have given the operator A a distinguished role. $V_c^{(\infty)}$ is constructed so as to contain only eigenstates of $A_{(1)}, A_{(2)}, \ldots$ corresponding to eigenvalue sequences $\{i\}$ selected from the countable set X of sequences with a finite number of ones. The other operators like B have eigenvalue sequences $\{j\}$ selected from the uncountable set Y of all sequences of zeros and ones. Despite this lack of symmetry between A and B we will obtain the results we want.

This situation is entirely familiar (see Appendix A). We want to construct a transformation from the basis $|a; \{i\}\rangle$ of normalized states to the basis $|b; \{j\}\rangle$ of infinite norm states with

$$\langle a; \{i\} | b; \{j\}\rangle = \prod_{r=1}^{\infty} \frac{\langle a, i_r | b, j_r\rangle}{\langle a, 0 | b, j_r\rangle}. \tag{TI}$$

(We can arbitrarily fix $\langle a; \{0\} | b; \{j\}\rangle = 1$.) This is achieved by the formulae

$$\langle a; \{i\} | \Psi\rangle = \int d\mu\{j\} \langle a; \{i\} | b; \{j\}\rangle \langle b; \{j\} | \Psi\rangle \tag{TII}$$

$$\langle b; \{j\} | \Psi\rangle \underset{(\mu)}{=} \sum_{\{i\}} \langle b; \{j\} | a; \{i\}\rangle \langle a; \{i\} | \Psi\rangle, \tag{TIII}$$

where $\langle b; \{j\} | a; \{i\} \rangle = \langle a; \{i\} | b; \{j\} \rangle^*$ are defined by (TI). The measure $d\mu\{j\}$ is defined over the set Y of sequences $\{j\}$. It occurs also in the formula (TIII), which is to be interpreted as saying that the sum on the right-hand side converges to the function of $\{j\}$ on the left-hand side in mean square with measure (μ), i.e.,

$$\int d\mu \{j\} \Big| \langle b; \{j\} | \Psi \rangle - \sum_{i_1, i_2, \ldots, i_N} \langle b; \{j\} | a; \{i_1, i_2, \ldots, i_N, 0, \ldots\} \rangle \times \langle a; \{i_1, i_2, \ldots, i_N, 0, \ldots\} | \Psi \rangle \Big|^2 \to 0 \quad \text{as} \quad N \to \infty. \qquad \text{(TIII')}$$

The crucial point is that the measure μ is *determined* by the requirement that the transformation be orthogonal. If we take $|\Psi\rangle = |a; \{i'\}\rangle$, then (TII) requires that

$$\prod_{r=1}^{\infty} \delta_{i_r i'_r} = \int d\mu\{j\} \prod_{r=1}^{\infty} \frac{\langle a, i_r | b, j_r \rangle \langle b, j_r | a, i'_r \rangle}{\langle a, 0 | b, j_r \rangle \langle b, j_r | a, 0 \rangle}.$$

Now take sequences with $i_r = i'_r = 0$, for $r > N$. Then

$$\prod_{r=1}^{N} \delta_{i_r i'_r} = \sum_{j_1 \cdots j_N} \prod_{r=1}^{N} \langle a, i_r | b, j_r \rangle \langle b, j_r | a, i'_r \rangle \frac{m(j_1 \cdots j_N)}{\prod_{r=1}^{N} p_{j_r}},$$

where

$$m(j_1 \cdots j_N) = \int d\mu\{j'\} \prod_{r=1}^{N} \delta_{j'_r j_r}$$

is the measure of all sequences beginning with $j_1, j_2, \ldots, j_N$. For $i_r = 0$, $r = 1, 2, \ldots, N$, there are 2^N independent equations in 2^N unknowns, whose only solution is

$$m(j_1 \cdots j_N) = \prod_{r=1}^{N} p_{j_r}.$$

Therefore the measure μ is precisely the probability measure that assigns independent probabilities p_0, p_1 to $j_r = 0, 1$.

To repeat in slightly different language, the transformation laws (TII), (TIII) set up an *isometry* between the Hilbert spaces l_2 of sequences $\langle a; \{i\} | \Psi \rangle$ with $\{i\} \in X$ and $L_2(\mu)$ of functions $\langle b; \{j\} | \Psi \rangle$ with $\{j\} \in Y$ and the measure μ on Y. Given that $\langle a; \{i\} | b; \{j\} \rangle$ satisfies (TI) the requirement of isometry determines μ.

VII. The Frequency Operator

We can now define a frequency operator corresponding to B by

$$F | b; \{j\} \rangle = f(\{j\}) | b; \{j\} \rangle,$$

where

$$f(\{j\})=\lim_{N\to\infty}\frac{1}{N}\sum_{r=1}^{N}j_r.$$

If the limit on the right-hand side does not exist, let $f(j\})=\frac{1}{2}$(upper limit + lower limit) $(1/N)\sum_{r=1}^{N}j_r$, which in fact could be used as the definition for all $\{j\}$. For a general state $|\Psi\rangle$

$$\langle b;\{j\}|\,F\,|\Psi\rangle=f(\{j\})\langle b;\{j\}|\Psi\rangle.$$

But the Strong Law of Large Numbers (SI) tells us that $f(\{j\})=p_1$ except for a set of measure zero when the measure μ is used on the set of sequences $\{j\}$. Hence

$$\langle b;\{j\}|\,F\,|\Psi\rangle=p_1\langle b;\{j\}|\Psi\rangle$$

except for a set of measure zero, i.e.,

$$F|\Psi\rangle=p_1|\Psi\rangle \qquad \text{for } any \text{ state } |\Psi\rangle\in V_c^{(\infty)};$$

i.e., $F=p_1 I$ as an operator equation. This is the desired result. Any state in the Hilbert space component $V_c^{(\infty)}$ is an eigenstate of the frequency operator F with eigenvalue p_1, where $p_1=|\langle b,1|a,0\rangle|^2$. A measurement of the proportion of states where the value of B is 1 will yield p_1. Postulate PIV′ implies the probability interpretation.

VIII. The Zero–One Law

In fact, a large class of interesting operators on $V_c^{(\infty)}$ are multiples of the identity: all symmetric functions of $B_{(1)}, B_{(2)}, \ldots$. (Note that this is not true for symmetric functions of $A_{(1)}, A_{(2)}, \ldots$; for example, $\sum_{r=1}^{\infty} A_{(r)}$.)

An operator G which is a symmetric function of $B_{(1)}, B_{(2)}, \ldots,$ satisfies $G|b;\{j\}\rangle=g(\{j\})|b;\{j\})\rangle$, where g is a symmetric function of $j_1, j_2, \ldots$. Consider the state $G|a;\{0\}\rangle$, which must have finite norm if $g\{j\}$ is bounded, i.e.,

$$\sum_{\{i\}}|\langle a;\{i\}|G|a;\{0\}\rangle|^2<\infty.$$

Recalling $\langle b;\{j\}|a;\{0\}\rangle=1$, we have

$$\langle a;\{i\}|G|\,a;\{0\}\rangle=\int d\mu\{j\}\,g(\{j\})\langle a;\{i\}|b;\{j\}\rangle$$

so that $\langle a;\{i\}|G|a;\{0\}\rangle$ is a symmetric function of $i_1, i_2, \ldots$ and $\omega(\{i\})=|\langle a;\{i\}|G|\,a;\{0\}\rangle|^2$ can only depend on the number of i_r that equal 1. But the number of sequences $\{i\}$ with n ones is infinite except for $n=0$, so $\omega(\{i\})=0$ except for $\{i\}=\{0\}$. So

$$G|a;\{0\}\rangle = g_0|a;\{0\}\rangle,$$

where g_0 is a constant. Hence for almost all $\{j\}$

$$\langle a;\{0\}|G|b;\{j\}\rangle = g_0\langle a;\{0\}|b;\{j\}\rangle;$$

i.e.,

$$g(\{j\}) = g_0$$

and therefore

$$G|\Psi\rangle = g_0|\Psi\rangle \qquad \text{for all} \quad |\Psi\rangle \in V_c^{(\infty)}.$$

Our Hilbert space setup has afforded an easy proof of the symmetric zero–one law of classical probability [7]. Take $g\{j\}$ to be the characteristic function of a set E of sequences $\{j\}$, (i.e., $g=1$ if $\{j\}$ is in E and $g=0$ if $\{j\}$ is not in E), with the property that if $\{j\}$ belongs to E then so do all sequences $\{j'\}$ obtained by finite permutations of $\{j\}$. We have shown that with our measure μ, $g\{j\}=g_0$ almost surely, where $g_0=$ zero or one. Hence E has probability zero or one.

IX. The Big Hilbert Space

We will now construct the non-separable Hilbert space $V^{(\infty)}$ which completely describes **S**, the system consisting of an infinite number of copies of S. The Hilbert space $V^{(\infty)}$ can be viewed as a direct sum of separable Hilbert spaces identical in structure to the $V_c^{(\infty)}$ we have already used,

Consider infinite sequences $\{\varphi\}=|\varphi_1\rangle, |\varphi_2\rangle, \ldots$ of unit vectors in V. Define an equivalence relation by

$$\{\varphi\} \sim \{\psi\} \qquad \text{iff} \quad \prod_{r>N} |\langle\psi_r|\varphi_r\rangle| > 0 \qquad \text{for some } N.$$

From each equivalence class $\mathscr{C}$ arbitrarily choose a sequence $\{\varphi\}$. (This requires the Axiom of Choice.) For each term $|\varphi_r\rangle$ in $\{\varphi\}$ choose a unit vector $|\varphi_r^\perp\rangle$ orthogonal to $|\varphi_r\rangle$. (Since V is two-dimensional, this amounts to choosing an arbitrary sequence of phases.) Now an orthonormal basis for $V_{\mathscr{C}}^{(\infty)}$ is given by the countable collection of sequences $\{i\}$ composed of 0's and 1's with finitely many 1's. A sequence $\{i\}$ corresponds to a sequence of the form $|\varphi_1\rangle, |\varphi_2^\perp\rangle, |\varphi_3\rangle, \ldots$, where $|\varphi_r\rangle$ or $|\varphi_r^\perp\rangle$ occurs in the rth place according as i_r is 0 or 1.

If $\{\psi\}$ is in $\mathscr{C}$, we can change the phases of the vectors $|\psi_r\rangle$ to make $\langle\psi_r|\phi_r\rangle$ real and non-negative for all r. The matrix element $\langle\{i\}|\{\psi\}\rangle=$

$\prod_{i_r=1} \langle \phi_r^\perp | \psi_r \rangle \prod_{i_r=0} \langle \phi_r | \psi_r \rangle$ is then well-defined for each $\{i\}$, and the sequence of states represented by $|\psi_1\rangle, |\psi_2\rangle, \ldots$, is represented in $V_{\mathscr{C}}^{(\infty)}$ by

$$\sum_{\{i\}} \langle \{i\} | \{\psi\} \rangle | \{i\} \rangle.$$

So, every sequence equivalent to $\{\varphi\}$ is represented up to irrelevant phases in $V_{\mathscr{C}}^{(\infty)}$, and every sequence is represented in $\oplus_{\mathscr{C}} V_{\mathscr{C}}^{(\infty)} = V^{(\infty)}$. $V^{(\infty)}$ is called the weak complete tensor product and $V_{\mathscr{C}}^{(\infty)}$ the weak incomplete tensor product [8].

We could repeat the change of basis in each component $V_{\mathscr{C}}^{(\infty)}$, obtaining measures $\mu_{\mathscr{C}}$. Each $\mu_{\mathscr{C}}$ has the property that $j_1, j_2, \ldots$ are independent, but now the probability that $j_r = 1$ is $|\langle \varphi_r | b, 1 \rangle|^2$. The eigenvalue of the frequency operator F in $V_{\mathscr{C}}^{(\infty)}$ is $\frac{1}{2}$(upper limit + lower limit)$(1/N) \sum_{r=1}^{N} |\langle \varphi_r | b, 1 \rangle|^2$. We could instead define upper and lower frequency operators, with eigenvalues equal to the upper and lower limits of $(1/N) \sum_{r=1}^{N} |\langle \varphi_r | b, 1 \rangle|^2$.

X. CONCLUSIONS

We have succeeded in proving the exact relation $F(\theta_i)|\psi\rangle^\infty = p_i |\psi\rangle^\infty$, where $|\psi\rangle^\infty$ is the vector representing an infinite number of copies of the single system in the state $|\psi\rangle$. (We have actually shown this for any $|\Psi\rangle$ which represents the single system state $|\psi\rangle$ in all but a finite number of copies and for certain linear combinations of these $|\Psi\rangle$'s.) The frequency operator $F(\theta_i)$ is the quantum operator associated with the measurement of the fraction of those systems which are found to yield the value θ_i when $\mathcal{O}$ is measured. The eigenvalue p_i is computed to be the single system inner product, $p_i = |\langle \theta_i | \psi \rangle|^2$.

To what extent is the system **S** of an infinite number of copies of S physically realizable and can the measurement of the frequency have physical meaning? This is like asking whether an infinite sequence of coin flips can be physically realized and what does the Strong Law of Large Numbers say in this case? Clearly a coin cannot actually be flipped an infinite number of times any more than someone can count to infinity; yet we can make sensible statements that refer to these operations. The Strong Law of Large Numbers says that if you did flip a coin an infinite number of times you would, with absolute certainty, find the proportion of heads to be one-half. Our quantum statement is that if you construct an infinite number of copies of S and measure the frequency of the outcome θ_i you would, with absolute certainty, find the frequency to be $p_i = |\langle \theta_i | \psi \rangle|^2$.

What is the state of the system, **S**, after a measurement of the frequency? [9]. In order to leave the state undisturbed by the measurement we must measure only the frequency and not each individual outcome. To see how this can be done, imagine that the single system is a spin one-half particle prepared with its spin along some arbitrary fixed direction. We wish to measure the fraction of systems whose spin is $+\frac{1}{2}$ along the z axis. For the moment let the number of systems S in **S** be a finite number, N. The operator $F=(1/N)\sum_{r=1}^{N}(S_{zr}+\frac{1}{2})$ measures the proportion of states with spin $+\frac{1}{2}$ along the z axis. It is possible to measure F without measuring each S_{zr}. Affix each of the particles to a lattice whose mass scales with N. Pass the lattice through a Stern–Gerlach analyzer oriented in the z direction. The deflection of the center of mass of the lattice, caused by magnetic field, depends only on the *total* spin in the z direction and in measuring the deflection you determine F, not the individual spins. Since the mass of the lattice scales with N, the maximum possible deflection is independent of N. We now imagine that we can discuss the case N equals infinity. The mass and force on the lattice are both infinite but the deflection is finite. A measurement of the deflection will yield, with absolute certainty, the value of the proportion equal to p calculated for the single system. The state of the total system will be unchanged by the measurement.

We have shown that the probability interpretation of quantum mechanics can arise if postulates PIV and PV are replaced by the weaker PIV′. The state $|\psi\rangle^{\infty}$ is an eigenstate of the frequency operator and a measurement of the frequency can be imagined which leaves the state unchanged. We are left with the question of what happens if we perform a measurement on a quantum system which is not in an eigenstate of the associated operator. What happens if we measure the z component of the spin of a single particle whose spin is along some direction other than the z axis? The orthodox answer is that once an outcome is established the state must become the eigenstate associated with the discovered eigenvalue. To allow for this we must reinstate postulate PV although PIV is still eliminated in favor of PIV′.

This line of reasoning assumes that the measuring device sits apart from the quantum system. The measuring act produces changes in the quantum states of the system which are not describable by the time evolution given by postulate PII. Postulate PV is needed as long as the measuring device is not part of the quantum system. However, you can consider a quantum system which contains both the thing being measured and the measuring device. A measuring device is then an interaction, described by a Hamiltonian, which induces correlations in the wave function of the total system between, say, macroscopic readouts and the spin states of a particle. The parts of the wave function corresponding to different readouts have no overlap and never interfere. This description of measurement, in which *everything* obeys the Schrödinger equation, requires a description of what a measuring device is but it can then be used to formulate quantum mechanics without postulate PV. This discussion of measurement is also consistent with the replacement of PIV by PIV′.

By working with postulate PIV′ instead of PIV we have formulated quantum mechanics without specific reference to the Hilbert space inner product and we have derived the probabilistic interpretation of the inner product. You may ask, Why begin with Hilbert space if you do not want the fundamental relation between two states $|\langle\psi_1|\psi_2\rangle|^2$ to have a direct physical interpretation? But Birkhoff and von Neumann [10] have shown how to axiomatize "quantum logic" without taking this relation between states as an input. The objects of their axiomatization are "experimental propositions." The axioms characterize the lattice of "propositions" sufficiently to force them to correspond to subspaces of a vector space over some field. The *negative* of an "experimental proposition," suitably axiomatized, is then shown to correspond to the *orthogonal* subspace defined with respect to an inner product. For our purpose we can take the selection of the complex number field to be empirical. The existence of the number $|\langle\psi_1|\psi_2\rangle|^2$ as a function of two states is then a consequence, and we have addressed the question of finding an interpretation.

Appendix A

The transformation laws (TI, TII, TIII) may be written in a much more familiar form for the special case when the matrix

$$\langle b; j \mid a; i\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

so that $p_0 = p_1 = \frac{1}{2}$. Label the sequence $\{j\}$ by the real number $x = \sum_{r=1}^{\infty} (j_r/2^r)$, thus identifying the space Y with the interval (0, 1) on the real line. For each sequence $\{i\}$, define the function

$$W_{\{i\}}(x) = \prod_{r=1}^{\infty} \frac{\langle b, j_r \mid a, i_r\rangle}{\langle b, j_r \mid a, 0\rangle} = \prod_{i_r=1} (-1)^{j_r}.$$

Another way of writing these functions is by defining a square wave function on the real line $S(x)$, by $S(x) = +1$, $2n < x < 2n+1$; $S(x) = -1$, $2n+1 < x < 2n+2$. Then

$$W_{\{i\}}(x) = \prod_{i_r=1} S(2^r x).$$

It is easy to see that the set of 2^n functions W with $i_r = 0$ for $r > n$ are linear combinations of the 2^n characteristic functions of the intervals $(k/2^n, (k+1)/2^n)$, and that the whole set of functions $W_{\{i\}}(x)$ is a complete orthonormal set on (0, 1) with the measure dx. (They are known as the Walsh functions [11].) Thus our transformation is precisely the transform on the interval (0, 1)

$$\psi(x) = \sum_{\{i\}} C_{\{i\}} W_{\{i\}}(x); \qquad C_{\{i\}} = \int_0^1 \psi(x)\, W_{\{i\}}(x)\, dx.$$

APPENDIX B

A more familiar example of how the measure μ is determined by the transformation function may be helpful. Consider the Hilbert space defined by an orthonormal basis $|n\rangle$, $n=0, 1, 2, \ldots$. Define operators a, $a^\dagger$, X by

$$a^\dagger|n\rangle=\sqrt{n+1}\,|n+1\rangle, \qquad a|n\rangle=\sqrt{n}\,|n-1\rangle$$

$$X=\frac{1}{\sqrt{2}}(a+a^\dagger).$$

If we seek an eigenstate of X,

$$Xf(a^\dagger)|0\rangle=xf(a^\dagger)|0\rangle$$

we find

$$\frac{df}{da^\dagger}+a^\dagger f=\sqrt{2}xf, \qquad f=\exp\left(\sqrt{2}\,xa^\dagger-\frac{1}{2}a^{\dagger 2}\right).$$

Of course the state $|x\rangle=f(a^\dagger)|0\rangle$ is not normalizable. To determine the measure on the space of functions $\langle x|\psi\rangle$, we require

$$\langle m|n\rangle=\int d\mu(x)\langle m|x\rangle\langle x|n\rangle.$$

An equivalent requirement is

$$\langle 0|e^{ua}e^{va^\dagger}|0\rangle=\int d\mu(x)\langle 0|e^{ua}|x\rangle\langle x|e^{va^\dagger}|0\rangle.$$

Simple calculations give

$$\langle 0|e^{ua}e^{va^\dagger}|0\rangle=e^{uv}$$

$$\begin{aligned}\langle 0|\,e^{ua}\,|x\rangle &= \langle 0|\,e^{ua}e^{\sqrt{2}xa^\dagger-a^{\dagger 2}/2}\,|0\rangle\\ &= \langle 0|e^{\sqrt{2}x(a^\dagger+u)-(a^\dagger+u)^2/2}\,|0\rangle\\ &= e^{\sqrt{2}xu-u^2/2}\end{aligned}$$

so we require

$$e^{uv}=\int d\mu(x)\,e^{\sqrt{2}xu-u^2/2+\sqrt{2}xv-v^2/2},$$

i.e.,

$$\int d\mu(x)\,e^{xz}=e^{z^2/4} \text{ for any complex } z.$$

This determines

$$d\mu(x)=\frac{1}{\sqrt{\pi}}e^{-x^2}\,dx.$$

We thus recover the x representation for the harmonic oscillator wave functions, except that we have chosen the normalization $\langle 0|x\rangle = 1$, and the Gaussian appears in the measure.

Appendix C

The generalization to the case when the single system S is described by an n-dimensional Hilbert space V is obvious, apart from one instructive difficulty. The transformation matrix on V, $\langle b, j|a, i\rangle$, is now an $n\times n$ unitary matrix $(i, j=0, 1, ..., n-1)$. Here $p_j = |\langle b, j|a, 0\rangle|^2$ satisfy $p_j \geqslant 0, \sum_j p_j = 1$. The Hilbert space $V_c^{(\infty)}$ is defined by the countable basis $|a; \{i\}\rangle$; $i_r = 0, 1, ..., m-1$; $i_r \neq 0$ for a *finite* set of r. The transformation on $V_c^{(\infty)}$, given by (TI, TII, TIII) works exactly as before, except for the special case when $p_j = 0$ for some values of j, say $j = n-m$, $n-m+1, ..., n-1$. (For $n=2$, this occurs only when $B=A$ or $B=I-A$.) We deal with this case as follows. Define the set Y to consist of all sequences $\{j\}$ with $j_r \geqslant n-m$ for a *finite* set E of r. Break up Y into the sum of a *countably* infinite collection of disjoint subsets Y_E consisting of sequences $\{j\}$ with $j_r \geqslant n-m$, $r\in E$; $j_r < n-m$, $r\notin E$. On each Y_E define a measure μ_E by assigning independent probabilities p_j to $j_r = j$, $e\notin E$, and assigning weight one to each value of $j_r \geqslant n-m$, $r\in E$. Thus the total measure of Y_E is $m^{d(E)}$, where $d(E)$ is the number of r in E. The transformation law (TI) is replaced by

$$\langle a; \{i\}|b; \{j\}\rangle = \prod_{r\in E} \langle a, i_r|b, j_r\rangle \prod_{r\notin E} \frac{\langle a, i_r|b, j_r\rangle}{\langle a, 0|b, j_r\rangle} \tag{TI*}$$

and in (TII), $\int_Y d\mu\{j\}$ is replaced by $\sum_E \int_{Y_E} d\mu_E\{j\}$. Note how this works for the case $n=2$, $B=A$, $\langle a, 0|b, 1\rangle = 0$. Each Y_E consists of a single sequence with $j_r = 1$, $r\in E$, $j_r = 0$, $r\notin E$. The set Y is identical with the set X of sequences $\{i\}$ and the transformaton laws reduce to $\langle a; \{i\}|\Psi\rangle = \langle b; \{i\}|\Psi\rangle$.

References

1. H. Feshbach and V. F. Weisskopf, Ask a foolish question ..., *Phys. Today* **41**, No. 10 (1988), 9.
2. H. Everett, III, *Rev. Mod. Phys.* **29** (1957), 454.
3. N. Graham, *in* "The Many Worlds Interpretation of Quantum Mechanics" (B. S. DeWitt and N. Graham, Eds.), p. 229, Princeton Univ. Press, Princeton, NJ, 1973.
4. J. B. Hartle, *Amer. J. Phys.* **36** (1968), 704.
5. A. Renyi, "Foundations of Probability," p. 197, Holden-Day, San Francisco, 1970.
6. W. Ochs, *J. Philos. Logic* **6** (1977), 473.
7. W. Feller, , "An Introduction to Probability Theory and Its Application," Vol. II. p. 122, Wiley, New York, 1966.
8. J. von Neumann, *Compositio Math.* **6** (1938), 1.
9. K. Gottfried, "Quantum Mechanics," Vol. I, Chap. IV, Benjamin/Cummings, New York, 1966.
10. G. Birkhoff and J. von Neumann, *Ann. Math.* **37** (1936), 823.
11. I. L. Walsh, *Amer. J. Math.* **55** (1923), 5.