International Journal of Theoretical Physics, Vol. 18, No. 3, 1979

An Application of Ideal Experiments to Quantum Mechanical Measurement Theory

Elihu Lubkin

Department of Physics, University of Wisconsin–Milwaukee, Milwaukee, Wisconsin 53201

Received January, 1969; revised January, 1979

It is shown that the rule for obtaining probabilities by squaring amplitudes is deducible from ideal experiments in a mechanics of unitary motion in a complex-linear space, with tensor product for making compound systems. Difficulty with tensor product in undoctored quaternionic quantum mechanics makes the argument inapplicable there. Except for the replacement by ideal experiments of a more formal unitary equivalence, the discussion is similar to that of Everett (1957). Diagonal expression of a general vector in the tensor product of two spaces is related to polar form of a matrix, in the Appendix.

1. INTRODUCTION

This article reviews the ordinary theory of measurement in quantum mechanics, but is new in that the rule that probabilities are absolute-squared amplitudes is deduced by arguments involving ideal experiments. The aim is to assume the mechanics, and to deduce the interpretation. *Therefore, a complex-linear space of states, motions unitary with respect to a Hermitean form, and tensor product for putting systems freely together are assumed.* Nevertheless, the arguments involving thought experiments make a further appeal to common sense, in a way which it would be inconvenient to state in advance in axiomatic form. The most formal device in these experiments is the convention that a system may be forced to execute any unitary transformation as a possible motion; I expose this axiomatically here because it would be unfair to pass it off as common sense. But, see Lubkin (1974). Complications which might arise from adapting the discussion, given for finite-dimensional inner-product spaces, to an infinite number of dimensions, are not discussed.

The probability interpretation implies unitarity of the motion and tensor

0020-7748/79/0300-0165$03.00/0

product for putting systems freely together. [Tensor product means that the states $i \to a_i$ and $j \to b_j$, wherein an a system has its state specified by amplitudes a_i for alternatives indexed by i, are combined by alternative-pairing and amplitude multiplication, into $(i, j) \to a_i b_j$. Thus, if the separate spaces have dimensions n_1, n_2, the linear space for the compound system spanned by the tensor products has dimension $n_1 n_2$. That the (i, j)th amplitude squares to the product of the probabilities $|a_i|^2$ and $|b_j|^2$ shows that it must be $a_i b_j$ up to a phase; to have this for all a-system orthonormal bases and b-system orthonormal bases removes the relative phase ambiguities.] This is to remind the reader how the mechanical principles of unitarity and tensor composition were discovered; but it is the converse proposition, that the mechanics alone leads to the squaring rule for probabilities, which is the point here.

The main argument is in Sections 2–9. Then come "Historical Remarks," etc. at the end, so as not to disturb continuity of argument.

2. ORTHOGONALITY OF STATES WHICH INEVITABLY INDUCE DIFFERENT RECORDS IN AN INSTRUMENT- AND REPLACEMENT OF THE INSTRUMENT BY A HERMITEAN OPERATOR (Wigner, 1952)

Let $x_i y_0 \to x_i y_i$ denote the evolution in time of a state x_i for an x system ("system of interest") together with an observer or instrument, a y system, from a time before the development of a record to a time after. (The notation $x_i' y_i$ would suggest a "recoil" of x_i in response to its interaction with the instrument, an inessential complication.) y_i is an instrument's state which bears the record that some macroscopic indicator reads i. Thus, x_i is a state of the system which inevitably produces the record i in a y instrument. If, moreover, instrument states y_i and y_j are grossly distinct, then the corresponding system states x_i and x_j are so different that they never produce the same result in a y measurement.

Then an argument of Wigner (1952) deduces that such x_i, x_j are orthogonal: The final states $x_i y_i$ and $x_j y_j$ are obviously orthogonal, owing to the gross ("macroscopic") difference in the instrument's states y_i, y_j, whence $x_i y_0$ and $x_j y_0$ were also orthogonal, owing to unitarity of the motion. Then $(x_i y_0, x_j y_0) = (x_i, x_j)(y_0, y_0) = 0$ gives $(x_i, x_j) = 0$.

The obvious orthogonality of $x_i y_i$ and $x_j y_j$ rests on the tensor-product combination of systems. Thus, if the instrument y has a subsystem, say, a pointer tip, which is obviously in mutually orthogonal parts of its state space on comparing y_i and y_j, one deduces that $(x_i y_i, x_j y_j) = 0$ without worrying about the remainder of the instrument or the x system, because (except for an inessential matter of linear combination) the inner product of a compound system is a product of inner products of tensor factors, and vanishes if any factor vanishes. *It must be agreed that the inner product is so chosen that grossly distinct states are orthogonal*; "macroscopic" need signify no more than this grossness.

Thus, associated to an instrument y, there is a family of mutually orthogonal subspaces of x space [subspaces: if $y_i = y_j$, then $(x_i + \alpha x_j)y_0 \to (x_i + \alpha x_j)y_i$], labeled by the distinct y outcomes. Let the range of x states be limited to the space generated by these sure-outcome-labeled subspaces, so that completeness is tautological.

If P_i is the projection on the ith x subspace, and if the distinct outcomes are labeled by distinct real numbers λ_i, then $Y = \sum \lambda_i P_i$ is the usual Hermitean operator acting on x space conveniently associated with the process of y observation. This is how I teach the assignment of a Hermitean operator to a process of observation.

3. THE APPEARANCE OF PROBABILITIES OTHER THAN 0 AND 1

The state $\sum a_i x_i y_0$ evolves into $\sum a_i x_i y_i$ by linearity of the motion. The idea that "the result" of a y observation on a nontrivial superposition of Y eigenstates will be seen in a simple sense by following the motion of the x-y complex, is frustrated by the production of a state with different gross attributes in different terms (von Neumann, 1932, 1955). This "ambiguity" is commonly taken to end the attempt to so use the motion; probabilities $|a_i|^2$ (with the x_i, y_i normalized) are adopted as a reflection of "ambiguity." This is correct, but too fast.

First, is a notion of probability really forced? That an observing y system will be forced to use probabilities is easy to see, if a y system is allowed to make a *run* of experiments on identically prepared states, instead of a single measurement.

To see this one *must* look at a run because it is the appearance of *different outcomes in a run* of identically prepared trials which constitutes the experience of nontrivial probability, defined as the ratio of "successes" to trials in a long run.

The "run" experiment, n trials, is

$$\sum a_{i_1} x_{i_1}^1 \times \cdots \times \sum a_{i_n} x_{i_n}^n y_{0,\ldots,0} \to \sum a_{i_1} \times \cdots \times a_{i_n} x_{i_1}^1 \times \cdots \times x_{i_n}^n y_{i_1,\ldots,i_n}$$

where $y_{0,\ldots,0}$ is the instrument ready to record the n outcomes, and $y_{i_1,\ldots,i_n}$ has recorded the outcomes, i_k in the kth trial. Among these there are observer's states with all i_k equal, that have not learned about the statistical aspect of the experiment, but there are others with i_k's *not all equal,* who must quote empirical probabilities other than 0 and 1 for some outcomes.

That states extreme as extreme points in the convex body of all mixtures need not be and here are not characterized by having only 0 and 1 as probability values, is discussed at length by von Neumann (1932, 1955). The point here is that in a discussion wherein states (which become known as pure states after we have the probabilities, so that von Neumann's analysis applies) are conceived of mechanically, the occurrence of probability values other than 0 and 1 is also understood mechanically. The "causal" propagation of wave functions itself forces included observers to experience nontrivial probabilities.

4. WHY RUNS ARE NOT USED FURTHER

The temptation is now strong to deduce the rule that $|a_i|^2$ is the ith outcome's probability, by examining the empirical probabilities: $y_{i_1,\ldots,i_n}$ has the subscript i appearing exactly $\sum_{k=1}^{n} \delta_{i,i_k}$ times, giving the unambiguous empirical probability $\sum_{k=1}^{n} \delta_{i,i_k}/n$. Each run has a definite empirical probability, which seems to leave no room for further theorizing. Unfortunately, the variety of empirical probabilities obtained, as the variety of runs $y_{i_1,\ldots,i_n}$ is all surveyed, is independent of the a_i's (excluding $a_i = 0$). Only by selecting a family of *typical* observers may the a_i's become relevant: some runs are more important than others. This requires a weighting of run-indexed observer's states by probabilities, for which one can posit the squared amplitudes $|a_{i_1,\ldots,i_n}|^2 = \prod_{k=1}^{n} |a_{i_k}|^2$, which, however, would put one logically back to zero. Perhaps one can yet use the run together with common-sense assumptions or formal axioms (Hartle, 1968; Finkelstein, 1963). I instead prefer to use the run only in the deduction above that observers experience probabilities and hence will demand some rule $i \to p_i$ for assigning probabilities to alternatives, and go to other arguments to fix the detailed rule.

The defeating point, in summary, is that the *list* of all 2^n possible runs for throwing a loaded (2-faced) coin n times contains no information about the loading, in spite of the fact that each run on this list does present a sharp opinion about the loading.

5. PSYCHOPHYSICAL PARALLELISM

If a z system records the result obtained by the y system, we have in obvious extension of the notation, $\sum a_i x_i y_0 z_0 \to \sum a_i x_i y_i z_0 \to \sum a_i x_i y_i z_i$. This final state is not a general state in xyz tensor-product space (this is unlike the case for 2 tensor factors; see Appendix), but no difficulty will arise therefrom.

By kinematically limiting the yz instrument's space to the space generated by the $y_i z_i$, excluding the $y_i z_j$ with $i \neq j$, the "general-state" form may be restored. The $y_i z_j$ with $i \neq j$ would represent incorrect transmission of the record.

Von Neumann's main argument (1932, 1955) is that the probabilities for the i-indexed alternatives are indifferent to whether the system's state is cut off from the instrument in the form $\sum a_i x_i$, or in the form $\sum a_i x_i y_i$. The proof is simply that $p_i = |a_i|^2$ for either [i.e., that $|(x_i, \sum a_j x_j)|^2 = |(x_i y_i, \sum a_j x_j y_j)|^2$, to give it a decently complicated appearance]. Though computationally trivial, it has significance. Two processes, law of motion of the wave function and reduction of wave function to probabilities, fit together smoothly: Whether the intermediate y part is part of the motion-treated state vector, or part of the nondescribed reducing instrument, the result is the same. Because the motion is an aspect of the theoretical dynamics ("physical") whereas the reduction is directed to the psychological reality of an observer's "result," this goes by the name of "psychophysical parallelism." The transmitting system y can be treated equivalently by either of the two mechanisms which are therefore satisfactorily parallel.

6. CASTING AWAY THE UNIT VECTORS; THE FIRST IDEAL EXPERIMENT

The probabilities $i \to p_i$ depend on the state vector $\sum a_i x_i$ and on the basis $(x_1, \ldots, x_n)$ separated by the measurement, and not on the instrument's states y_i in $\sum a_i x_i y_i$, by the principle of psychophysical parallelism. Thus, the p_i's depend only on the list of amplitudes and orthogonal unit vectors $(a_1, \ldots, a_n; x_1, \ldots, x_n)$.

The following argument concludes that p_i is the same for $(a_1, \ldots, a_n; x'_1, \ldots, x'_n)$, where the x'_i are any other orthogonal unit vectors, and therefore only on $(a_1, \ldots, a_n)$.

Let y be an instrument which separates the x_i and also separates a list of orthonormal vectors u_i in a space orthogonal to that generated by the x_i, but which does not distinguish between x_i and u_i. Consider the scheme

$$\sum a_i x_i y_0 \xrightarrow{U} \sum a_i u_i y_0 \xrightarrow{V} \sum a_i u_i y_i$$

$$\sum a_i x_i y_0 \xrightarrow{V} \sum a_i x_i y_i \xrightarrow{U} \sum a_i u_i y_i$$

V is the motion by which measurement is achieved; U is a motion which rotates the x basis into the u basis, $Ux_i = u_i$. Since the same final state is attained, $i \to p_i$ must be the same. In the first case, the data are $(a_1, \ldots, a_n; u_1, \ldots, u_n)$; in the second, before the U motion, they are $(a_1, \ldots, a_n; x_1, \ldots, x_n)$. The U motion does not modify the results, because $x_i \to u_i$ does not change the y property.

Now consider the same argument, where, however, an x' basis is U' rotated to the same u basis, and a y' instrument separates the x_i', the u_i, but not x_i' from u_i; the scheme is

$$\sum a_i x_i' y_0' \xrightarrow{U'} \sum a_i u_i y_0' \xrightarrow{V'} \sum a_i u_i y_i'$$

$$\sum a_i x_i' y_0' \xrightarrow{V'} \sum a_i x_i' y_i' \xrightarrow{U'} \sum a_i u_i y_i'$$

The conclusion is that $(a_1, \ldots, a_n; u_1, \ldots, u_n)$ yields the same p_i's as $(a_1, \ldots, a_n; x_1', \ldots, x_n')$.

Consequently, the $(a_1, \ldots, a_n; x_1, \ldots, x_n)$ and the $(a_1, \ldots, a_n; x_1', \ldots, x_n')$ yield the same p_i's.

The u_i were used to avoid difficulty which would otherwise arise from possible linear dependence of $(x_1, \ldots, x_n, x_1', \ldots, x_n')$.

7. CASTING AWAY THE PHASES

It follows as a corollary that the p_i's are the same for $(|a_1|, \ldots, |a_n|)$ as for $(a_1, \ldots, a_n)$: From $\sum a_i x_i y_i = \sum |a_i| x_i' y_i$, where $x_i' = a_i |a_i|^{-1} x_i$ (zero terms are omitted), either of (x_i) or (x_i') is a separation basis for the y measurement; also, $\sum a_i x_i = \sum |a_i| x_i'$. Therefore the same p_i's follow from $(a_1, \ldots, a_n; x_1, \ldots, x_n)$ or from $(|a_1|, \ldots, |a_n|; x_1', \ldots, x_n')$, and from the result of the last section, the x_i or x_i' may be omitted.

Therefore, $p_i = f_i(|a_1|, \ldots, |a_n|)$. By relabeling, the general result may be had from $p_1 = f_1(|a_1|, \ldots |a_n|) = F(|a_1|, \ldots, |a_n|)$.

The following argument shows that F does not actually depend on the $|a_2|, \ldots, |a_n|$.

8. CASTING AWAY THE OTHER AMPLITUDES

Let y be a crude instrument, which separates x_1 (recording state y_1) from the space generated by $x_2, \ldots, x_n$, but which does not separate these from each other (recording state $y_{\geqslant 2}$).

Consider the schemes

$$\sum a_i x_i y_0 \xrightarrow{U} \left(a_1 x_1 + \sum_{i \geqslant 2} a_i' x_i\right) y_0 \xrightarrow{V} a_1 x_1 y_1 + \sum_{i \geqslant 2} a_i' x_i y_{\geqslant 2}$$

$$\sum a_i x_i y_0 \xrightarrow{V} a_1 x_1 y_1 + \sum_{i \geqslant 2} a_i x_i y_{\geqslant 2} \xrightarrow{U} a_1 x_1 y_1 + \sum_{i \geqslant 2} a_i' x_i y_{\geqslant 2}$$

V is the motion which eventuates in measurement, U a general unitary transformation motion in the $x_{\geqslant 2}$ space orthogonal to x_1, which leaves x_1 fixed. The freedom of choice of U is expressed by arbitrariness of the a_i', subject to

$$\sum_{i \geqslant 2} |a_i'|^2 = \sum_{i \geqslant 2} |a_i|^2 = 1 - |a_1|^2$$

As before, identity of the final states guarantees that p_1 is the same in both schemes; in the second one, it is established when the amplitudes are $(a_1, a_2, \ldots, a_n)$, but is unaffected by the subsequent U motion, the result 1 and the state x_1 being unchanged; the change of state for the result $\geqslant 2$ being such as not to affect the result.

By choosing $a_2' = (1 - |a_1|^2)^{1/2}$, $a_3' = \cdots = a_n' = 0$, one obtains

$$p_1 = F(|a_1|, (1 - |a_1|^2)^{1/2}, 0, \ldots, 0) \equiv f(|a_1|^2)$$

which is explicitly a function of $|a_1|$, or equivalently of $b_1 = |a_1|^2$, only.

Since a fine measurement may be made subsequent to the coarse sorting into bins 1 and $\geqslant 2$ by separating the bin $\geqslant 2$, without affecting the ratio (1-outcomes)/(trials), the p_1 for the above coarse measurement must apply also to a fine measurement.

Therefore, there is a universal function f of one variable such that

$$p_i = f(b_i), \qquad \text{where } b_i = |a_i|^2$$

9. THE ABSOLUTE-SQUARE RULE

To show that $f(b) = b$: Construct a state (using, if necessary, a suitable freely presumed unitary motion, starting from a state with $b_1 = 1$), with $b_1 = b_2 = \cdots = b_n = 1/n$. Since each $p_i = f(1/n)$ and $\sum_{i=1}^{n} p_i = 1$, we have $\sum_{i=1}^{n} f(1/n) = 1$, or $f(1/n) = 1/n$. Hence $f(b) = b$ is proved for numbers of form $1/n$, with n a positive integer.

Next, suppose $b + \sum 1/n = 1$, $0 \leqslant b \leqslant 1$, where the sum is of some finite number of fractions of form $1/n$. Build a state with $b_1 = b$, using for the other b_i the fractions in that finite sum: we will have a unit vector because these numbers add to 1. Then $p_1 = f(b)$, and $\sum_{i\geqslant 2} p_i = \sum f(1/n) = \sum 1/n = 1 - b$. But also, $\sum_{i\geqslant 2} p_i = 1 - p_i$, whence $p_1 = b$; again $f(b) = b$.

The set of b for which $f(b) = b$ is proved now includes the numbers between 0 and 1 of form $1 - \sum 1/n$. Among these are all the terminating binary fractions, a dense set; hence continuity of f would imply $f(b) = b$ in general, concluding the argument. It is reasonable to presume continuity, because the amplitudes of a state cannot practically be determined with precision. A discontinuous rule from the amplitudes could not fit stable empirical application of the rule.

10. LANDAU TRACING

Since the measurement process is now reduced to the usual rule $p_i(x) = |a_i|^2$, $a_i = (x_i, x)$, for the probability of the ith outcome from state x, we have in the usual way the formula $\mathrm{Tr}\, PY$ for expectation values of observables Y in vector states, and then in mixtures thereof by convex combination. (As previously noted, the vector states are then seen to be pure in the sense of extreme points among the mixtures, in the usual way.)

If now the projection density matrix $P_{\sum a_i x_i y_i}$ for the pure state $\sum a_i x_i y_i$ is used in conjunction with observables A which act only on the x system, $A = A' \otimes 1$, in the usual way, the tracing out of the y indices provides a reduced density matrix P' and an expectation-value formula Tr $P'A'$ with all quantities referent to the x system only, where P' is the diagonal matrix on the (x_i) basis with diagonal elements $|a_i|^2$. Amplitudes a_i have been replaced by probabilities $|a_i|^2$ by *explicitly* ignoring the observer y. Ignoring the y observer was not automatic here and required tracing out because we are here pretending not to be ourselves the y observer. In this sense, the mechanism of Landau (1927) wherein confinement of attention to a subsystem yields a reduced density matrix with extra entropy, although derived from measurement theory, then encompasses measurement theory. Were the density-matrix formulas regarded self-evident, this would even be an independent derivation of the probability interpretation! But even if the probability interpretation is attained otherwise, as for example in the preceding sections, Landau tracing teaches us something. Namely, Landau's tracing out of superfluous indices applied to the state-vector treatment of measurement by motion demonstrates that the quality of the $|a_i|^2$ as probabilities emerges computationally from the violence done to the state function by an observer's necessary self-ignoration.

The ugliness of the probabilities is a necessary reflection of the ugliness of exploring a state by looking into the memories of computers described along with other junk within the state. A temporal sequence would seem also to be suspect as a necessary artifact of this mode of analysis (and an artifact associated with increasing entropy). It seems fortunate that the crude nature of backwards-memory exploration of dynamics could become evident in a context in which the dynamical entity itself, the state, evolves naively in "time." Wanted: a dynamics without time, and therefore in all probability, without space.

11. DIFFICULTY WITH QUATERNIONIC QUANTUM MECHANICS

In the lattice-of-tests theory (Birkhoff and von Neumann, 1936; Jauch, 1968[1]), possible systems admitted by all the axioms include real and quater-

[1] This emphasizes the lattice-theoretic viewpoint, and has numerous pertinent references. The peculiar omission however of references to the many-worlds interpretation is mirrored in a degree of awkwardness in the discussion of some of the "paradoxes." Further similar material might be extant from the *Lattice Theory and General Quantum Mechanics Symposium*, 26–28 December 1968, University of Missouri–Rolla, Rolla, Missouri.

nionic quantum mechanics (Finkelstein et al., 1962) as well as the usual complex quantum mechanics. The arguments here are unchanged if the a_i are real, but not if they are quaternions.

Starting with the rule that $p_i = |a_i|^2$, with quaternionic amplitudes a_i calculated by inner products with orthonormal quaternionic unit vectors, defined $(x, y) \equiv \sum \overline{x_i} y_i$, one does succeed in writing $\sum p_i \lambda_i$, λ_i real, as (x, Hx), with $(x, Hy) = (Hx, y)$, and $(Hx)_i = \sum H_{ij} x_j$, given by a quaternionic matrix acting on the left, and right-linear (i.e., $H(x + yq) = Hx + (Hy)q$, for q a right quaternionic "scalar" factor). Also $(x, yq) = (x, y)q$, and $(xq, y) = \bar{q}(x, y)$, but $(x, qy) = (\bar{q}x, y)$ only; a *left* scalar would not come *out* of an inner product.

However, there is no obvious notation for tensor product; using $(i, j) \rightarrow x_i y_j$ allows for a pathological x left linearity, and a y right linearity; $(i, j, k) \rightarrow x_i y_j z_k$ leaves a y factor with no linearity. Going from system to subsystem by "ignoring" in a 2-system density matrix is stopped by (x, Hx) failing to turn into a density-matrix formula; $\sum \overline{x_i} H_{ij} x_j \neq \sum x_j \overline{x_i} H_{ij}$, owing to the non-commutativity of the quaternions. Difficulty in forming tensor product is noted in Finkelstein et al. (1962), but it is not clear that this difficulty is overcome there.

The idea that an observer be alternatively a subsystem is part of psycho-physical parallelism, of the Wigner argument, and of the program of finding a mechanical underpinning for the interpretation.

The arguments presented above for the statistical interpretation of complex quantum mechanics, therefore do not extend to quaternionic quantum mechanics, owing to the character of quaternionic quantum mechanics in being an axiomatic confrontation between states and tests not adequately equipped with a mechanism for compounding systems.

12. HISTORICAL REMARKS

Although my meager reading will not support a historical survey (Jauch, 1968; Thomas, 1958; Ludwig, 1961; Wigner, 1963; de Witt and Graham, 1973), I append this section and the next in order to orient my own remarks with respect to other work.

A fantastic insight appears in Lucretius ($\sim$55 B.C.; Latham, 1951). Since a symmetric initial state may not evolve neatly and causally into a random world, Lucretius suggests slight departures from causality. Dirac (1938–1939) also notes that quantum jumps may suffice to produce the randomness for thermodynamic equilibration. That Landau tracing is jumpy enough: Lubkin (1978).

The Copenhagen school and "viewpoint" are too well known to require comment.

The finishing touch to this work is von Neumann's remarks on measurement theory per se and on its relationship to thermodynamics (1932, 1955) where he credits Landau (1927) with the discovery that the density matrix appropriate to a subsystem usually has nonzero entropy even if the larger system has none. This suggests that the Lucretian randomness is not fundamental, but is an artifact of looking at things incompletely, a point that is clinched by the logical impossibility of inclusion of an instrument or of himself by an observer whose aim it is to consider that instrument only incidentally.

Readers of von Neumann's book have frequently and independently understood this message, in spite of his warning (von Neumann, 1955, pp. 419–420) that a mind must interpret a physical system in terms of probability-weighted results obtained by an axiomatically independent process. The derivative character of the probabilities and a review of the von Neumann work was given in sharply didactic form by Everett (1957) and Wheeler (1957).

The point of the present paper is that I prefer to replace a completely formal usage of unitarity in Everett's work by the unitarity of the motion in ideal experiments.

Everett emphasizes the tree of alternatives which is developed by measurements. This is basic to understanding Einstein–Podolsky–Rosen effects (Einstein, Podolsky, and Rosen, 1935; Dicke and Wittke, 1960), and generally in understanding von Neumann. It appears as the "cat-in-the-box" experiment (Schrödinger, 1935; Jauch, 1968) of the "Copenhagen viewpoint." Truncation of the tree to uncorrelated branches is the entropy-increasing process of Landau. Yet ever since people worried about spreading wave packets, there is a feeling that the dilution of the wave function in a tree cannot somehow fit with the apparent substantiality of the world. Although an observer looking back into his memory is obviously never bothered about his own unsubstantiality ["I think, therefore I am" (Descartes, 1637)], I find it an additional if superfluous comfort to note that a more numerical substantiality is embodied in the conservation laws, and that these laws have no reference to the treelike complexity which may be discerned in a wave function: substantiality is unworried by trees; each branch has all the baryons.

13. RELATION TO THE LATTICE-THEORETIC ("EXISTENTIAL") APPROACH

Von Neumann has done the best, in a certain existential sense, with lattice of tests, and states (assignations of probabilities to tests) over it (Birkhoff and von Neumann, 1936; Jauch, 1968): unpredictability tempered by probabilities is the primitive experience (Descartes, 1637); therefore a work like this present one which looks into a mathematical "machine" first in order to later find observers in the machine and ask about their memories, is in this existential view nonoptimal.

Two excuses:

1. It is nice to see the probabilities deduced, nevertheless; to see how the statistical character may be viewed as the fault of making contact with a nicely operating machine by the ugly approach of gathering up memory sequences in recording machines: I think, therefore I distort.

2. Even existential physics builds a physical model, which then does hold up the physical end of psychophysical parallelism. Thus, even if the existential approach is superior in its minimality, it *generates* a mechanism, and seeks the observers in the mechanism. There is a circle here. Where one breaks into a circle is to some extent arbitrary. Breaking in at different points involves different kinds of axioms, however, and therefore breaking in variously allows for more opportunities to modify current theory. Thus, the lattice-state approach contains axioms which are obvious from the nature of observation, e.g., the existence of an opposite to every test (interchange the labels "success" and "failure"), but also axioms (e.g., the "modular" axiom) whose only excuse is a superficial crispness of form ("clarity"), and that they work for the classical and finite-dimensional quantum mechanical cases: von Neumann himself invites (Birkhoff and von Neumann, 1936, Section 18; von Neumann, 1955, p. 309) the reader to criticize these weak axioms, and hopes to find more systems in the lattice-state mode of reasoning by doing so (motivation for his work on "continuous geometries"). Starting elsewhere on the psychophysical circle may profitably enlarge the scope for creating modifications.

> Imagine, for example, a mechanical approach wherein the observers are found only by making an approximation. It may be very difficult to reason backwards to certain important, simple features of the machine from the existential data of such observers; i.e., nice underlying features may turn into ugly, small effects in the observers' data, and may therefore elude discovery from the lattice-state approach.

Caveat: In the existential approach, purity of a state is its property of being an extreme point. "Mechanical" may be defined as the preservation of such purity within the general motion of mixtures. It is not obvious that laws of motion, mechanical in this sense, are necessary (Thomas, 1958; Ludwig, 1961; Wigner, 1963). Starting with a "machine" and then finding the observers does *not* beg this question, as the way the observers see the "machine" as revealed by their memory records need not be "mechanical" in the sense of preservation of purity by what *they* describe as "motion."

Of course, in the main discussion here, the particular assumptions of complex linearity, unitarity of motion, and tensor-product composition fit only the usual quantum mechanics, and is "mechanical" in both senses.

We perversely feel that we understand something only when that understanding involves an approximation. Linearity in quantum mechanics suffers from absoluteness. If only we could feel that linearity applies because something is small! How departure from linear unitarity affects the statistical "facts" seen by internal observers is an issue which would therefore make us feel better even about the linear unitary case. Such thoughts are close to recent work of Bodgan Mielnick (1977 or 1978).

APPENDIX. RELATION OF DIAGONAL FORM FOR A TENSOR-PRODUCT SPACE VECTOR TO POLAR FORM OF A MATRIX

Von Neumann (1932, 1955) introduces the operation $(y_j, \sum a_i x_i y_i) \to a_j x_j$ of partial inner product, essentially Everett's (1957) "relative-state" operation, in proving that the form $\sum a_i x_i y_i$ of a "measurement state" is not formally distinct from the form $\sum a_{ij} x_i y_j$ of a general state in the tensor-product space. (The state space of the y system is severely restricted here; e.g., broken states of the instrument are not admitted.) This is nice to know, to prevent pointless work on bringing out nonexistent special attributes of measurement, but is not essential.

The following reference to polar form may be useful in locating this item in the usual operator-theoretic paraphernalia:

To show that there exist unitary transformations $x_i' = \sum u_{ik} x_k$ and $y_j' = \sum v_{jm} y_m$ such that $\sum a_{ij} x_i y_j = \sum a_i' x_i' y_i'$, note that the former is

$$\sum a_{ij}(u^{-1})_{ik}(v^{-1})_{jm} x_k' y_m' = ({}^t u^{-1})_{ki} a_{ij} (v^{-1})_{jm} x_k' y_m'$$

Whence the problem is equivalent to finding unitary matrices U, V so that UAV is diagonal. By writing $A = PW$ in polar form, P Hermitean non-negative, and W unitary, and then unitarily diagonalizing P, $P = U^{-1}A'U$, A' diagonal non-negative, one obtains $A = U^{-1}A'UW$; or $A' = UAV$, with $V = W^{-1}U^{-1}$. If A is not originally a square matrix, make it square by adding zero rows or columns, to simplify the grammar.

ACKNOWLEDGMENTS

Professor L. E. Parker told me of Graham's work, of Hartle's similar work, and kindly provided a preprint of a lecture of de Witt's. Professor H. A. Stapp showed me the reference to Dirac. My knowledge of Bogdan Mielnick's work comes from a preprint entitled *Mobility* he sent me.

REFERENCES

Birkhoff, G., and von Neumann, J. (1936). *Annals of Mathematics*, **37**, 823.
Descartes, R. (1637). *Discours de la Methode*, 2e partie.
de Witt, B. C., and Graham, R. N. (1973). *The Many Worlds Interpretation of Quantum Mechanics*, Princeton University Press, Princeton, New Jersey.
Dicke, R. H., and Wittke, J. P. (1960). *Introduction to Quantum Mechanics*, Addison-Wesley, Reading, Massachusetts.
Dirac, P. A. M. (1938–1939). *Proceedings of the Royal Society of Edinburgh*, **59**, 2, 122.
Einstein, A., Podolsky, B., and Rosen, N. (1935). *Physical Review*, **47**, 777.
Everett, H., III. (1957). *Reviews of Modern Physics*, **29**, 454.
Finkelstein, D., Jauch, J. M., Schiminovich, S., and Speiser, D. (1962). *Journal of Mathematical Physics*, **3**, 207.
Finkelstein, D. (1963). *Transactions of the New York Academy of Sciences*, **25**, 621–637.
Hartle, J. B. (1968). *American Journal of Physics*, **36**, 704.
Jauch, J. M. (1968). *Foundations of Quantum Mechanics*, Addison-Wesley, Reading, Massachusetts.
Landau, L. D. (1927). *Zeitschrift für Physik*, **45**, 430.
Latham, R. E. (1951). *On the Nature of the Universe*, p. 66, a translation of Lucretius (~55 B.C.), Penguin, Toronto.
Lubkin, E. (1974), *Journal of Mathematical Physics*, **15**, 673–674.
Lubkin, E. (1978). *Journal of Mathematical Physics*, **19**, 1028–1031.
Titus Lucretius Carus (~55 B.C.). *De Rerum Natura*. See Latham (1951).
Ludwig, G. (1961). In *Werner Heisenberg und die Physik unserer Zeit*, Vieweg u. Sohn, Braunschweig.
Mielnick, B. (1977 or 1978). *Mobility*, preprint.
Schrödinger, E. (1935). *Naturwissenschaften*, **48**, 52.
Thomas, L. H. (1958). *Physical Reviews*, **112**, 2129.
von Neumann, J. (1932, 1955). *Mathematical Foundations of Quantum Mechanics*, tr. Beyer, R. T., Princeton, 1955.
Wheeler, J. A. (1957). *Reviews of Modern Physics*, **29**, 454.
Wigner, E. P. (1952). *Zeitschrift für Physik*, **133**, 101, footnote 2, p. 102.
Wigner, E. P. (1963). *American Journal of Physics*, **31**, 6.