[2]K. R. Symon, *Mechanics*, 3rd ed. (Addison-Wesley, Reading, MA, 1971).

[3]Pierre Louis Moreau de Maupertuis, *Essai de Cosmologie*, 1759, quoted by W. Yourgrau and S. Mandelstam in *Variational Principles in Dynamics and Quantum Mechanics*, 3rd ed. (Dover, New York, 1968), p. 20.

[4]See, for example, Max Born and Emil Wolf, *Principles of Optics*, 6th ed. (Pergamon, Oxford, 1980), pp. 101–132.

[5]Born and Wolf (Ref. 4), pp. 738–746, provide a standard treatment from the optical point of view. A standard treatment on the mechanical side is that of H. Goldstein, *Classical Mechanics*, 2nd ed. (Addison-Wesley, Reading, MA, 1980), pp. 484–492. See also Yourgrau and Mandelstam (Ref. 3), pp. 58–64.

[6]Three recent examples are provided by the following. R. J. Black and A. Ankiewicz, "Fiber-optic analogies with mechanics," Am. J. Phys. **53**, 554–563 (1985). J. W. Blaker and M. A. Tavel, "The application of Noether's theorem to optical systems," Am. J. Phys. **42**, 857–861 (1974). W. B. Joyce, "Classical particle description of photons and phonons," Phys. Rev. D **9**, 3234–3256 (1974).

[7]This point has also been stressed by J. A. Arnaud, "Analogy between optical rays and nonrelativistic particle trajectories: A comment," Am. J. Phys. **44**, 1067–1069 (1976).

[8]We have derived Snell's law using the constancy of the component of dr/da parallel to the boundary, i.e., we have used $(dx/da)_1 = (dx/da)_2$. It is worth remembering that, although these quantities play the roles of velocities, they are not really velocities. The actual x component of the velocity is $v_x = (dx/da)(da/dt) = (c/n^2)dx/da$. The constancy of dx/da across the boundary therefore implies $v_{x1} n_1^2 = v_{x2} n_2^2$, so that the horizontal velocities are definitely not the same in the two media.

[9]In this problem we have supposed the *index of refraction* to vary linearly with the height above the road; as shown, the ray is a catenary. Alternatively, one may take the *speed of light* to vary linearly with the height. The results then are that the ray is a circular arc, with the center located at the height, where the speed of light goes to zero. (Our thanks to F. Danes for pointing this out.) This alternative problem is most easily solved by starting from the conservation of "energy," i.e., the equation that is the last entry in Table I.

[10]Born and Wolf (Ref. 4).

[11]M. V. Klein, *Optics* (Wiley, New York, 1970).

[12]E. W. Marchand, *Gradient Index Optics* (Academic, New York, 1978).

[13]The question we have posed is, more strictly, this: For what function $n(r)$ will there exist a circular orbit centered on the origin *for every r*. The condition we obtained, $rdn = -ndr$, is satisfied for all r only by the function $n = k/r$. Only in this particular case, then, do such circular orbits exist for all radii. There are, however, an infinity of possible functions $n(r)$ for which the condition $rdn = -ndr$ is satisfied *at one or more particular values for r*. In such a case, a circular orbit centered on the orgin will be possible, but only at particular, isolated radii. An example of such a case is provided by the Maxwell "fish-eye," i.e., by the function $n(r) = n_0[1 + (r/b)^2]^{-1}$. As is well known, the general light orbit in this system is a circle whose center is displaced from the origin. The off-centeredness of the circular orbit depends upon n_0 and b, as well as upon the initial conditions. [Born and Wolf (Ref. 4), pp. 147–149.] However, for the one particular case $r = b$, the off-centeredness vanishes. Indeed, it may be verified by direct calculation that for the Maxwell fish-eye, $dn/dr = -n/r$ only at $r = b$.

[14]R. P. Luneberg, *Mathematical Theory of Optics*, Brown University mimeographed notes, 1944 (University of California, Berkely, CA, 1964).

[15]This problem provides an instance in which the optical "potential energy," $-n^2/2$, does not correspond to the usual choice in mechanics. We have $U = r^2/(2r_0^2) - 1$ in the optical case, while the usual choice in mechanics is $kr^2/2$. This constant shift in the scale of the "potential energy" does not, of course, alter the trajectories. However, our usual convention of taking $-n^2/2$ as the analog of the potential energy results, as usual, in a "total energy" of zero, thus producing a paradox—how can there be motion in the case of the harmonic oscillator if the "total energy" is zero? The resolution of the paradox simply involves the choice of the zero of "potential energy." If the optical "potential energy" were defined exactly as is customary in the mechanical case, the "total energy" would be 1.

Probability theory in quantum mechanics

L. E. Ballentine
Department of Physics Simon Fraser University, Burnaby, British Columbia, Canada V5A 1S6

(Received 12 August 1985; accepted for publication 27 September 1985)

The abstract theory of probability and its interpretation are briefly reviewed, and it is explicitly demonstrated that the formalism of quantum mechanics satisfies the axioms of probability theory. This refutes the suggestions which have occasionally been made that "classical" probability theory does not apply to quantum mechanics. Several erroneous applications of probability theory to quantum mechanics are examined, and the nature of the errors are exposed. It is urged that more attention be given to probability theory in the physics curriculum.

I. INTRODUCTION

It is generally agreed that probability must be employed at a fundamental level in the interpretation of quantum mechanics. Yet the concept and theory of probability are usually treated very loosely and superficially. I have not seen any textbook which demonstrates just how the axioms of probability theory are satisfied by the formalism of quantum mechanics. The first objective of this paper is to remedy that shortcoming, and in order to do so I first give a brief outline of the theory of probability, and those aspects of its interpretation that are relevant to this task.

It has occasionally been claimed that "classical" probability theory does not apply to quantum mechanics. Those claims are sometimes based on misinterpretations of quantum mechanics, but more often on misinterpretations of probability theory. Some of those erroneous claims are examined in the latter part of this paper.

II. AXIOMATIC PROBABILITY THEORY

The mathematical content of the theory of probability concerns the properties of a function $P(A|B)$, which may be read as "the probability of A conditional on B." A and B may be "events," in which case $P(A|B)$ is "the probability that event A will happen under the conditions specified by the occurrence of event B." Alternatively, A and B may be propositions, in which case $P(A|B)$ is "the probability that A is true given that B is true." These alternative interpretations, as well as the interpretation of "probability" will be discussed in the next section. But all such interpretations can be based upon a common mathematical formalism, which derives from a set of axioms. For the sake of definiteness I will use the language of events in this section.

It is desirable to treat sets of events as well as elementary events. Therefore we introduce notations for certain composite events: $\sim A$ ("not A") denotes the nonoccurrence of A; $A \cdot B$ ("A and B") denotes the occurrence of both A and B; $A \vee B$ ("A or B") denotes the occurrence of at least one of the events A or B. For brevity these sets of events will also be referred to as "events." The three operators $(\sim, \cdot, \vee)$ are called *negation*, *conjunction*, and *disjunction*. In evaluation of complex expressions the negation operator has the highest precedence. Thus $\sim A \cdot B = (\sim A) \cdot B$, and $\sim A \vee B = \sim A) \vee B$.

Several different but mathematically equivalent forms of the axioms can be given.[1] The particular choice used here is influenced by the work of Cox[2]:

$$0 \leqslant P(A|B) \leqslant 1, \tag{1}$$

$$P(A|A) = 1, \tag{2}$$

$$P(\sim A|B) = 1 - P(A|B), \tag{3}$$

$$P(A \cdot B|C) = P(B|A \cdot C)P(A|C). \tag{4}$$

Axiom (2) states the convention that the probability of a certainty (the occurrence of A given the occurrence of A) is one, and (1) says that no probabilities are greater than the probability of a certainty. Axiom (3) expresses the intuitive notion that the probability of nonoccurrence of an event increases as the probability of its occurrence decreases. It also implies $P(\sim A|A) = 0$, that is to say, an impossible event (the nonoccurrence of A given that A occurs) has zero probability. Axiom (4) states that the probability that two events both occur (under some condition C) is equal to the probability of occurrence of one of the events multiplied by the probability of the second events, given that the first event has occurred. (In the work of Cox[1] these quantitative axioms are derived from more fundamentatl qualitative postulates.)

The probabilities of negation $(\sim A)$ and conjunction $(A \cdot B)$ of events each required an axiom. However, no further axioms are required to treat disjunction because $A \vee B = \sim(\sim A \cdot \sim B)$; in word, "$A$ or B" is equivalent to the negation of "neither A nor B." Thus from (3) we have

$$P(A \vee B|C) = 1 - P(\sim A \cdot \sim B|C), \tag{5}$$

which can be evaluated from the existing axioms to yield (see Appendix A)

$$P(A \vee B|C) = P(A|C) + P(B|C) - P(A \cdot B|C). \tag{6}$$

If $P(A \cdot B|C) = 0$ we say that A and B are mutully exclusive on condition C. Then (6) reduces to the rule of *addition of probabilities for mutually exclusive events*,

$$P(A \vee B|C) = P(A|C) + P(B|C). \tag{7}$$

This is often used as an axiom instead of (3). Indeed it is possible to derive (3) from (7) (see Appendix B). We thus have two equivalent sets of axioms: (1)–(4) or (1), (2), (4), and (7). The former set of axioms is more elegant because it applies to all events, whereas (7) applies only if A and B are mutually exclusive. Nevertheless the latter set is commonly used, and it has advantages in certain situations.

A very important concept in probability and its applications is that of *independence* of events. If $P(B|A \cdot C) = P(B|C)$, that is to say, if the occurrence of A has no influence on the probability of B, then we say that B is independent of A (under condition C). From (4) we then obtain

$$P(A \cdot B|C) = P(B|C)P(A|C). \tag{8}$$

The symmetry for this formula implies that *independence* is a mutual relationship; if B is independent of A then also A is independent of B. This notion is called *statistical* or *stochastic independence* in order to distinguish it from other notions such as causal independence.

III. INTERPRETATION OF PROBABILITY CONCEPTS

The abstract probability theory, consisting of axioms, definitions, and theorems, must be supplemented by an *interpretation* of the term "probability." This provides the correspondence rule by means of which the abstract theory can be applied to practical problems. There are many different interpretations of probability because anything that satisfies the axioms may be regarded as a kind of probability.

One of the oldest interpretations is the *limit frequency* interpretation. If the conditioning event C can lead to either A or $\sim A$, and if in n repetitions of such a situation the event A occurs m times, then it is asserted that $P(A|C) = \lim_{n\to\infty} (m/n)$. This provides not only an interpretation, but also a definition of probability in terms of a numerical frequency ratio. The axioms of the abstract theory can be derived as theorems of the frequency theory. In spite of its superficial appeal, the limit frequency interpretation has been widely discarded, primarily because there is no assurance that the above limit really exists for the actual sequences to which one wishes to apply probability theory.

The defects of the limit frequency interpretation are avoided without losing its attractive feature (close contact with observable data) in the *propensity* interpretation.[3] The probability $P(A|C)$ is interpreted as a measure of the tendency, or propensity, of the physical conditions described by C to produce the result A. It differs mathematically from the older limit frequency theory in that "probability" remains a fundamental undefined term, and is not redefined or derived from anything more fundamental. However its relationship to frequency emerges, suitably qualified, in a theorem (the law of large numbers). It differs conceptually from the frequency theory in viewing probability (propensity) as a characteristic of the physical situation C that may potentially give rise to a sequence of events, rather that a property (frequency) of an actual sequence of events. This fact is emphasized by always writing probability in the conditional form $P(A|C)$, and never merely as $P(A)$. The propensity interpretation is particularly well suited for application to quantum mechanics.

In addition to the probability P, one must choose an interpretation of the arguments A and C of the function $P(A|C)$. So far we have spoken of them as being *events*, but one can also treat them as *propositions*. In many cases the difference between the two interpretations is merely verbal. Corresponding to the event A is the proposition "event A has occurred." But one can usefully consider propositions that do not correspond to events. For example one can consider the probability (conditional on specified experimental evidence) that the electronic charge is within one part in a thousand of its conventional published value. This is the point of view that is used in *inductive inference*, and we shall have some occasion to apply it. But in spite of the seemingly greater generality of the interpretation of the arguments as propositions, it is often more convenient in quantum mechanics to interpret them as events. For example, let A be the proposition "the position of the particle lies between q_1 and q_2." Let B be the proposition "the momentum of the particle lies between p_1 and p_2." Quantum mechanics provides a means of computing the probability that A is true, and also for computing the probability that B is true. But it does not provide any formula for the probability that the compound proposition $A \cdot B$, (A and B), is true. Whether or not it is a defect of the present formulation of quantum mechanics, this limitation can be reasonably accommodated within the event interpretation. In computing the probability $P(A|C)$ of an event A, one must specify all the physical conditions C which are relevant. This may reasonably be held to include the configuration of any measuring apparatus, since it can influence the outcome of an event. Since different apparatuses are used to measure position than to measure momentum, one will be dealing with $P(A|C_q)$ and $P(B|C_p)$, where C_q includes the configuration of the position measuring device and C_p includes the configuration of the momentum measuring device. But one has no occasion to consider events A (detection of position within a certain range) and B detection of momentum within a certain range) under a common condition C, and so one does not need to compute $P(A \cdot B|C)$ in this case.

IV. PROBABILITY AND FREQUENCY

Although no direct connection between frquency and probability is postulated in the propensity interpretation, a close connection emerges through a theorem known as the *law of large numbers*. The derivation of the simplest form of this theorem is outlined below.

Let X be some quantity which, under some condition C, may take on a range of non-negative values, with $P(x<X<x+dx|C)=g(x)dx$. Then for any $\epsilon>0$ the probabilistic average of X (denoted $\langle X\rangle$) satisfies

$$\langle X\rangle = \int_0^\infty g(x)x\,dx$$

$$\geqslant \int_\epsilon^\infty g(x)x\,dx \geqslant \epsilon \int_\epsilon^\infty g(x)dx = \epsilon P(X\geqslant\epsilon|C).$$

Thus we have $P(X\geqslant\epsilon|C)\leqslant\langle X\rangle/\epsilon$, which is known as *Chebyshev's* inequality. We may apply this inequality to nonnegative variable $|X-\langle X\rangle|^2$, instead of X, obtaining

$$P(|X-\langle X\rangle|\geqslant\epsilon|C)\leqslant\langle|X-\langle X\rangle|^2\rangle/\epsilon^2. \qquad (9)$$

Now let us consider an experiment E which may have outcome A, with probability $P(A|E)=p$. In a sequence of n identical repetitions of E (denoted E^n) the event A may occur m times, $(0\leqslant m\leqslant n)$. We refer to $f=m/n$ as the frequency of outcome A in a realization of the experimental sequence E^n. One expects, intuitively, that f should be close to p as n becomes large. The following theorem justifies and makes more precise this intuitive expectation. Let K_i have the value 1 if the outcome of the ith repetition of E is A, otherwise K_i is zero. The frequency of A is given by $f=\Sigma_{i=1}K_i/n$. It is not difficult to show that $\langle f\rangle=p$. Substitution of f for X in (9) then yields

$$P(|f-p|\geqslant\epsilon|E^n)\leqslant\left\langle\left[\sum_{i=1}^n(K_i-\langle K_i\rangle)\right]^2\right\rangle\Big/(n\epsilon)^2$$

$$=\sum_i\sum_j[\langle(K_i-p)(K_j-p)\rangle]/(n\epsilon)^2. \qquad (10)$$

Since the various repetitions of E are independent, we have

$\langle(K_i-p)(K_j-p)\rangle=\langle(K_i-p)\rangle\langle(K_j-p)\rangle$ for $i\neq j$.

Thus (10) becomes

$$P(|f-p|\geqslant\epsilon|E^n)\leqslant\langle(K_i-p)^2\rangle/n\epsilon^2, \qquad (11)$$

the average on the right-hand side being independent of i. This result, which is an instance of the *law of large numbers*, asserts that the probability of the frequency of A being more than ϵ away from p converges to zero as n becomes infinite. From the practical point of view, this is the most important theorem in probability theory, establishing the connection between abstract probabilities and frequencies in observable data.

It should be noted, in passing, that the full proof of (11) uses axiom (4) only to the extent that it is needed to derive the addition rule (7). So if, as is sometimes done, we were to choose (1), (2), (4), and (7) as axioms, instead of (1)–(4), we could say that axiom (4) was not needed to derive the law of large numbers.

V. PROBABILITY IN QUANTUM MECHANICS

In quantum mechanics a dynamical variable R is represented by a self-adjoint operator $\mathbf{R}$, whose eigenvalues are the possible values of R:

$$\mathbf{R}|r_n\rangle=r_n|r_n\rangle. \qquad (12)$$

According to a standard postulate of quantum mechanics, the probability of obtaining the particular value $R=r_n$ is given by

$$P(R=r_n|\Psi)=|\langle r_n|\Psi\rangle|^2, \qquad (13)$$

in the simplest case of a discrete nondegenerate eigenvalue spectrum and a pure state represented by the vector Ψ. (All vectors here are assumed to have unit norms.) But it is not sufficient to merely assert that certain mathematical expressions are probabilities unless it can be shown that they satisfy the mathematical theory of probability. In particular, we must verify that such expressions obey the four axioms of probability theory and, in appropriate circumstances, the independence property (8).

The expression on the left-hand side of (13) can be read as "the probability that the dynamical variable R has the value r_n conditional on Ψ." The latter portion of this statement requires some comment because the state vector Ψ is not a physical object. Its significance is twofold. Firstly, it is an abstract mathematical object from which the probability distributions of observable quantities can be calculated. Secondly, to assert that the state vector is Ψ can be regarded as implying that the system has undergone a corresponding state preparation procedure, which could be described in more detail but all of the relevant information is contained in the specification of Ψ.

It is clear that (13) satisfies axiom (1), this being a direct consequence of the Schwartz inequality. If the state vector is the eigenvector $|\Psi\rangle = |r_n\rangle$, then (13) becomes $P(R = r_n | r_n) = 1$, which is the equivalent of axiom (2). Axiom (3) follows from the more general additivity rule (7). The events described by $R = r_1$, $R = r_2$, etc. are mutually exclusive, and the additivity rule, $P[(R = r_1) \vee (R = r_2) | \Psi] = |\langle r_1 | \Psi \rangle|^2 + |\langle r_2 | \Psi \rangle|^2$, holds almost by definition. We shall defer consideration of axiom (4) because it can be better treated by a more general formalism. If the system consists of two independent noninteracting components, which are only formally regarded as a single system, then the state vector can be written in the form $|\Psi\rangle = |\psi\rangle^{(a)} |\psi\rangle^{(b)}$, where the superscripts refer to the two components. The joint probability distribution for the dynamical variables $R^{(a)}$ and $R^{(b)}$, belonging to components a and b, respectively, is

$$P(R^{(a)} = r_m, R^{(b)} = r_n | \Psi) = |\langle r_m | \psi \rangle^{(a)}|^2 |\langle r_n | \psi \rangle^{(b)}|^2, \tag{14}$$

in full agreement with (8). We have now verified that the quantum mechanical postulate (13) satisfies all the ingredients of probability theory that are needed to derive the law of large numbers. Indeed if we interpreted the components a and b above as systems in the sequence of measurements E^n, decribed in Sec. IV, we could recapitulate the derivation of the law of large numbers in the language of quantum mechanics.

The most general description in quantum mechanics is by means of the *state operator* $\boldsymbol{\rho}$, which has unit trace, is self-adjoint, and is non-negative definite:

$$\mathrm{Tr}\, \boldsymbol{\rho} = 1, \tag{15}$$

$$\boldsymbol{\rho} = \boldsymbol{\rho}^\dagger, \tag{16}$$

$$\langle u | \boldsymbol{\rho} | u \rangle \geqslant 0 \text{ for all vectors } u. \tag{17}$$

In the special case of a pure state, represented above by the vector $|\Psi\rangle$, the state operator is $\boldsymbol{\rho} = |\Psi\rangle\langle\Psi|$. Associated with the any dynamical variable R is a family of projection operators $\mathbf{M}_R(\Delta)$ which are related to the eigenvalues and eigenvectors of (12) as follows:

$$\mathbf{M}_R(\Delta) = \sum_{r_n \in \Delta} |r_n\rangle\langle r_n|, \tag{18}$$

where the sum is over all eigenvectors (possibly degenerate) whose eigenvalues lie in the subset Δ. The probability that the value of R will lie within Δ is postulated to be

$$P(R \in \Delta | \boldsymbol{\rho}) = \mathrm{Tr}[\boldsymbol{\rho} \mathbf{M}_R(\Delta)]. \tag{19}$$

It is easily verified that the general form (19) reduces to (13) in the appropriate special case.

We must now verify that (19) satisfies the axioms of probability theory. Axiom (1) follows directly from (15) and the fact that $\mathbf{M}_R(\Delta)$ is a projection operator. The analogue of (2) is obtained if we chose a state prepared in such a manner that the value of R is guaranteed to lie within Δ. This will be so for those states which satisfy $\boldsymbol{\rho} = \mathbf{M}_R(\Delta)\boldsymbol{\rho}\mathbf{M}_R(\Delta)$, for which (19) is identically equal to 1. Axiom (3) follows from the additivity rule (7). To verify it we consider two disjoint sets Δ_1 and Δ_2, the union of which is denoted $\Delta_1 \cup \Delta_2$. Now $(R \in \Delta_1) \vee (R \in \Delta_2)$ is equivalent to $R \in (\Delta_1 \cup \Delta_2)$. Since the two sets Δ_1 and Δ_2 are disjointed it follows that $\mathbf{M}_R(\Delta_1)\mathbf{M}_R(\Delta_2) = 0$, and the projection operator corresponding to the union of the sets is just the sum of the separate projection operators, $\mathbf{M}_R(\Delta_1 \cup \Delta_2) = \mathbf{M}_R(\Delta_1) + \mathbf{M}_R(\Delta_2)$. Hence it is clear that (7) is satisfied. The factorization property (8) follows, as did (14), from the factorization of the state function for two independent uncorrelated systems, $\boldsymbol{\rho} = \boldsymbol{\rho}^{(a)} \otimes \boldsymbol{\rho}^{(b)}$. We have now verified that the general statistical postulate (19) satisfies all of those parts of probability theory that are needed to derive the key theorem, the law of large numbers.

The remaining axiom (4) is not essential for most of the applications of probability in quantum mechanics [provided that (7) replaces (3) as an axiom], but is must be considered for completeness. Let R and S be two dynamical variables, represented by the operators $\mathbf{R}$ and $\mathbf{S}$ whose eigenvalues and eigenvectors are given $\mathbf{R}|r_n\rangle = r_n|r_n\rangle$ and $\mathbf{S}|s_n\rangle = s_n|s_n\rangle$. The corresponding projection operators are denoted $M_R(\Delta_a)$ and $M_S(\Delta_b)$. Finally, let A denote the event of R taking on a value within the set Δ_a, and let B denote the event of S taking on a value within the set Δ_b. We must now evaluate each of the three probabilities in (4) with the conditional event C being the preparation of a general state represented by $\boldsymbol{\rho}$.

The joint probability on the left-hand side of (4), $P(A \cdot B | \boldsymbol{\rho})$, can be evaluated from the formalism of quantum mechanics only if the operators $\mathbf{R}$ and $\mathbf{S}$ are commutative (the corresponding projection operators $\mathbf{M}_R$ and $\mathbf{M}_S$ then also being commutative). In that case the product $\mathbf{M}_R(\Delta_a)\mathbf{M}_s(\Delta_b)$ is also a projection operator, and the desired joint probability is given by (19) to be $P(A \cdot B | \boldsymbol{\rho}) = \mathrm{Tr}[\boldsymbol{\rho}\mathbf{M}_R(\Delta_a)\mathbf{M}_S(\Delta_b)]$. But there is no accepted formula in quantum mechanics for a joint probability distribution for dynamical variables whose operators do not commute.

On the right-hand side of (4), the second factor $P(A|\boldsymbol{\rho})$ is given directly by (19) with $\Delta = \Delta_a$. However the first factor $P(B | A \cdot \boldsymbol{\rho})$ requires careful interpretation. The second argument of the probability function, which we have called "the conditional event," must describe the actual physical conditions to which the probability (or "propensity") refers. It does *not* denote mere subjective information or personal belief (as would be the case in a subjective interpretation of probability). Therefore, just as $\boldsymbol{\rho}$ signifies that the system has undergone a certain state preparation, so $A \cdot \boldsymbol{\rho}$ implies that is has been subjected to additional filtering interactions that ensure the value of R lies Δ_a. In principle one should analyze the dynamics of this process in detail in order to compute the resulting state function. But if this filtering process does nothing but remove unacceptable values of R then it is reasonable to represent its result by the projected and renormalized state operator, $\boldsymbol{\rho}' = \mathbf{M}_R(\Delta_a)\boldsymbol{\rho}\mathbf{M}_R(\Delta_a)/\mathrm{Tr}[\mathbf{M}_R(\Delta_a)\boldsymbol{\rho}\mathbf{M}_R(\Delta_a)]$. One then obtains the following result for the right -hand side of (4):

$$P(B | A \cdot \boldsymbol{\rho}) P(A | \boldsymbol{\rho}) = \mathrm{Tr}[\boldsymbol{\rho}'\mathbf{M}_S(\Delta_b)]\mathrm{Tr}[\boldsymbol{\rho}\mathbf{M}_R(\Delta_a)]$$
$$= \mathrm{Tr}[\mathbf{M}_R(\Delta_a)\boldsymbol{\rho}\mathbf{M}_R(\Delta_a)\mathbf{M}_S(\Delta_b)].$$

If $\mathbf{M}_R$ commutes with $\mathbf{M}_S$, then this expression furthur reduces to

$$P(B | A \cdot \boldsymbol{\rho}) P(A | \boldsymbol{\rho}) = \mathrm{Tr}[\boldsymbol{\rho}\mathbf{M}_R(\Delta_a)\mathbf{M}_S(\Delta_b)]$$
$$= P(A \cdot B | \boldsymbol{\rho}),$$

in agreement with axiom (4). Thus we see that this last axiom of probability theory is obeyed by the formalism of quantum mechanics provided the probability of the joint event $A \cdot B$ is defined in the formalism. The restriction, in this case, to dynamical variables whose operators commute is not a restriction on the applicability of "classical" probability theory to quantum mechanics. It is rather a limitation of the formalism of quantum mechanics, in that it does not assign meaning to the conjunction of arbitrary events. (See the discussion at the end of Sec. III.)

VI. ERRONEOUS APPLICATIONS OF PROBABILITY THEORY IN QM

A. The double slit

This example has been repeated many times in slightly differing versions. The first of which may be that due to Feynman.[4] The experiment consists of a particle source, a screen with two slits (labeled no. 1 and no. 2) in it, and a dectector. By moving the detector and measuring the particle count rate at various positions, one can measure the probability of a particle passing through the slit system and arriving at the point X. If only slit no. 1 is open the probability of detection at X is $P_1(X)$. If only slit no. 2 is open the probability of detection at X is $P_2(X)$. If both slits are open the probability of detection is $P_{12}(X)$. Now passage through slit no. 1 and passage through slit no. 2 are certainly exclusive events, so one might expect, from (7), that $P_{12}(X)$ should be equal to $P_1(X) + P_2(X)$. But experiment clearly shows that this is not true, hence it might be concluded that the rule (7) of probability theory does not hold in quantum mechanics.

In fact the above argument draws its radical conclusion from an incorrect application of probability theory. One is well advised to beware of probability statements expressed in the form $P(X)$ instead of $P(X|C)$. The second argument may be safely omitted only if the conditional event or information is clear from the context, and is *constant* throughout the problem. This is not the case in the double slit example. The probability of detection at X in the first case (only slit no. 1 open) should be written as $P(X|C_1)$, where the conditional information C_1 includes (at least) the state function Ψ_1 for the particle beam and the screen state S_1 (only slit no. 1 open). In the second case (only slit no. 2 open) the probability should be written as $P(X|C_2)$, where C_2 includes the state function Ψ_2 for the particle beam and the screen state S_2 (only slit no. 2 open). In the third case (both slits open) the probability is of the form $P(X|C_3)$, where C_3 includes the state function Ψ_{12} (approximately equal to $\Psi_1 + \Psi_2$, but this fact plays no role in our argument) and the screen state S_3 (both slits open). We observe from experiment that $P(X|C_3) \neq P(X|C_1) + P(X|C_2)$. This fact, however, has no bearing on the validity of rule (7) of probability theory. Essentially this counter argument to Feynman was given by Koopman.[5]

B. The superposition fallacy

The following argument is taken from Sec. 2.2 of the textbook by Trigg,[6] although other versions of it exist.
"Classical probabilities are compounded according to the relation

$$P(B'|A') = \sum_{C'} P(B'|C')P(C'|A'), \qquad (20?)$$

where the summation is over all members C' of a set of nonoverlapping states connecting A' and B'."
It is then noted that in quantum mechanics this relation is satisfied by amplitudes, $\langle B'|A'\rangle = \Sigma_{C'} \langle B'|C'\rangle\langle C'|A'\rangle$. Since the probabilities are the squares of the amplitudes. $P(B'|A') = |\langle B'|A'\rangle|^2$, it follows that the equation (20?) of the "classical theory" can hold only if the quantum interference terms are negligible.

There is no doubt that (20?) fails to hold as written, but we must examine more closely its status with respect to probability theory. We may presume from the context that A', B', and C' denote events which can be characterized by unique values for certain corresponding dynamical variables, and moreover that the set of possible values of C', say, is a mutually exclusive and exhaustive set. That is to say, no more than one such value can occur at a time (exclusive), and there are no other possible outcomes in the relevant class of events than one of the values from this set (exhaustive). Put yet another way, if the set $\{C_1, C_2, C_3 \cdots\}$ contains all possible values of C' then the disjunction of all those possibilities, $C_1 \vee C_2 \vee C_3 \vee \cdots$, is a certainty. In attempting to derive (20?) we make use of (4) to obtain $P(B' \cdot C'|A') = P(B'|C' \cdot A')P(C'|A')$, and then sum over all possible values of C'. Since each of $B' \cdot C_1$, $B' \cdot C_2, \cdots$ are exclusive, it follows from (7) that $\Sigma_{C'} P(B' \cdot C'|A') = P[B' \cdot (C_1 \vee C_2 \vee C_3 \vee \cdots)|A']$ $= P(B'|A')$. Therefore the correct deduction from "classical" probability theory is

$$P(B'|A') = \Sigma_{C'} P(B'|C' \cdot A')P(C'|A'), \qquad (21)$$

rather than the questionable (20?).

It is now apparent that the quantum mechanical superposition principle for amplitudes is in no way incompatible with the formalism of probability theory, and that the contrary claim was based on an incorrect application of probability theory. The error in this example is very similar to that in Sec. VI A. In the former case the conditional argument of the probability function was omitted, leading to an erroneous conclusion. In this case only a part of the relevent conditional information was included by writing $P(B'|C')$ instead of $P(B'|C' \cdot A')$ in (20?). That would be permissible only if it could be shown that the additional information was not relevant, which is evidently not the case.

C. The reciprocity fallacy

This example is taken from Sec. 2.3 of the same textbook by Trigg.[6]
"If the times involved in the specification of the two states are the same, the probability $P(B'|A')$ (*here I alter Trigg's notation slightly in order to conform to that used in this paper*), is actually the probability that the system satisfy two conditions simultaneously. This cannot depend on the order in which the conditions are stated, so we require

$$P(B'|A') = P(A'|B')\text{''}. \qquad (22?)$$

A probability theorist will immediately recognize that the author of the above quotation has confused the *conditional* probability $P(B'|A')$ with a *joint* probability. "The probability that the system satisfy two conditions simultaneously" (under some unspecified prior condition C), should be denoted as $P(A' \cdot B'|C)$, which has nothing to do with (22?).

The spurious equation (22?) draws its superficial plausibility from the amplitude relation, $P(B'|A') = |\langle B'|A'\rangle|^2$, which appears to support (22?), though not, of course, the specious argument that preceded it. But even that apparent connection is misleading.

To see more clearly the subtlety which is involved, let us rewrite the relation between probability and amplitude in a more explicit form,

$$P(R = r_n | \Psi) = |\langle r_n | \Psi \rangle|^2. \quad (13)$$

This is interpreted as the probability that the dynamical variable R has the value r_n, conditional on the state being Ψ. But what about the inverse probability, $P(\Psi | R = r_n)$, which is the probability that the state was Ψ on the condition that R was found to have the value r_n? Does it have the same value (13)?

One can easily show, by means of a simple example, that the inverse probability does not have the value (13). Suppose that space is divided into cells and that R is a discrete position variable. If Ψ is chosen to be a wavefunction localized within the nth cell, then one will have $P(R = r_n | \Psi) = 1$. From a knowledge of Ψ one can predict with certainty that the particle lies in the nth cell. But suppose, on the other hand, that Ψ is unknown and one has only the single measurement result $R = r_n$. Then all that one can infer about Ψ is that it must have been nonzero in the nth cell. There is no assurance that the (definite but unknown) state preparation procedure which led to the state Ψ would, if repeated, yield the same value for R, and so there is no reason to believe Ψ to be localized. Therefore $P(\Psi | R = r_n)$ will definitely be less than one.

The relation between the direct and inverse probabilities can be deduced by observing that since the left-hand side of (4) is symmetrical in A and B, so must be the right-hand side. Hence we obtain

$$P(A | B \cdot C) = P(B | A \cdot C) P(A | C) / P(B | C), \quad (23)$$

which is known as *Bayes theorem*. Applied to the above example, it yields

$$P(\Psi | R = r_n \cdot C) = P(R = r_n | \Psi \cdot C) P(\Psi | C) / P(R = r_n | C). \quad (24)$$

Here C denotes any other relevant prior information about Ψ, and it may be ignored if there is none.

At this point we have left the domain of events and the propensity interpretation of probability. This is so because the occurrence of the state Ψ is not an event which can be causally influenced by the subsequent determination that $R = r_n$. We are instead engaging in inductive inference, attempting to infer what the state might have been, on the basis of information which is not adequate to determine it uniquely. The value of this first factor on the right-hand side of (24) is simply $P(R = r_n | \Psi \cdot C) = P(R = r_n | \Psi) = |\langle r_n | \Psi \rangle|^2$, since no further information C about Ψ is relevant if Ψ is given. The last factor, $P(R = r_n | C)$, is called the *prior* probability that $R = r_n$. It might be a uniform distribution over the portion of space that can possibly be occupied by the particle. The second factor, $P(\Psi | C)$, is the *prior* probability, conditional on whatever information C may be available, that the state should be Ψ. It is rather difficult to evaluate, and must be regarded as only a degree of reasonable belief. It should now be apparent that we are very unlikely to obtain $P(\Psi | R = r_n) = P(R = r_n | \Psi)$, contrary to the spurious equation (22?).

VII. CONCLUSIONS

In this paper I have briefly reviewed axiomatic probability theory and its interpretation, with emphasis on those aspects that are most relevant to quantum mechanics. By demonstrating explicity that the axioms of probability theory are satisfied by the formalism of quantum mechanics, I have refuted any and all claims that "classical" probability theory is not valid in quantum mechanics. The only anomaly is the fact that joint probability distributions for two or more dynamical variables are not conventionaly defined unless the corresponding operators are commutative. But quantum mechanics can hardly be said to contradict (literally, "speak against"), probability theory on this point, since the accepted formalism of quantum mechanics is simply silent here.

Some examples of erroneous applications of probability theory in quantum mechanics have been exposed and analyzed. In view of the fact that these errors were committed by well educated physicists, one is led to the conclusion that probability theory needs greater emphasis in our curriculum. This is especially so in relation to quantum mechanics, where probability enters at a fundamental level. I hope that this paper will be helpful in achieving this goal.

APPENDIX A: DERIVATION OF (6), (BASED ON SEC. 5 OF COX[2])

To evaluate (5) by means of axioms (1)–(4) we require a lemma:

$$\begin{aligned}(P(X \cdot Y | C) | P(X \cdot \sim Y | C) &= P(X | C) P(Y | X \cdot C) + P(X | C) P(\sim Y | X \cdot C) \\ &= P(X | C) \{ P(Y | X \cdot C) + P(\sim Y | X \cdot C) \} \\ &= P(X | C). \end{aligned} \quad (A1)$$

Here we have used (4) and (3). Using (A1) with $X = \sim A$ and $Y = \sim B$, we obtain

$$P(\sim A \cdot \sim B | C) = P(\sim A | C) - P(\sim A \cdot B | C).$$

In the first term we use now (3), and in the second term we use (A1) with $X = B$, $Y = A$. This yields

$$P(\sim A \cdot \sim B | C) = 1 - P(A | C) - [P(B | C) - P(B \cdot A | C)].$$

Upon substitution of this result into (5) we obtain (6).

APPENDIX B: DERIVATION OF AXIOM (3) FROM (7)

By noting that A and $\sim A$ are mutually exclusive, and substituting $B = \sim A$ in (7), we obtain $P(A \vee \sim A | C) = P(A | C) + P(\sim A | C)$. Now $A \vee \sim A$, ("A or not A"), is intutively a certainty, and if we set its probability equal to 1 then we immediately obtain (3). The only gap in this proof lies in the fact that *certainty* is defined by axiom (2), and we should relate our intuitive notion of certainty to that definition. A formal proof that $P(A \vee \sim A | C) = 1$ is to be found on p. 17 of the book by Cox.[2]

[1]T. L. Fine, *Theories of Probability, and Examination of Foundations* (Academic, New York, 1973).

[2]R. T. Cox, *The Algebra of Probable Inference* (Johns Hopkins, Baltimore, MD, 1961).

[3]K. R. Popper, "The Propensity Interpretation of Calculus of Probability, and the Quantum Theory," pp. 65–70 in *Observation and Interpreta-*

tion, edited by S. Korner (Butterworths, London, 1957).
[4]R. P. Feynman, "The Concept of Probability in Quantum Mechanics," pp. 533–541 in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (University of California, Berkeley, CA, 1951).
[5]B. O. Koopman, "Quantum Theory and the Foundations of Probability," pp. 97–102 in *Applied Probability*, edited by L. A. MacColl (McGraw-Hill, New York, 1955).
[6]G. L. Trigg, *Quantum Mechanics* (Van Nostrand, Princeton, NJ, 1964).

Untenability of simple ensemble interpretations of quantum measurement probabilities

Daniel T. Gillespie
Research Department, Naval Weapons Center, China Lake, California 93555

(Received 22 July 1985; accepted for publication 25 October 1985)

Some interpretations of quantum mechanics try to describe the probabilistic nature of quantum measurements in a way that allows a system in a pure state to always have precise (if not simultaneously knowable) values for all its observable variables. In what might be called a *simple ensemble interpretation*, the quantum mechanical state vector is assumed to define a statistical ensemble of identically prepared systems, each of which has precise values for all its observable variables, and the act of measurement is equivalenced to a straightforward sampling of that ensemble. It is shown here, in a manner that should be suitable for a first course in quantum mechanics, that a simple ensemble interpretation is not possible for some quantum systems, and hence is untenable as a general premise. The analysis is essentially a repackaging of the Bell argument against hidden variables, although hidden variables are never invoked here.

I. INTRODUCTION

A generally accepted tenet of quantum mechanics is the *measurement prediction postulate*[1]: If an observable A with operator **A** is measured on a system whose state vector is $|\Psi\rangle$, then the probability of obtaining the eigenvalue a_i of **A** is equal to the square modulus of the projection $|\Psi\rangle$ onto the subspace spanned by the eigenvectors of **A** belonging to a_i.

Proponents of the Copenhagen interpretation of quantum mechanics use the measurement prediction postulate in conjunction with the *measurement reduction postulate*.[1] According to the latter, if the eigenvalue a_i is actually obtained in a measurement of A on the system in the state $|\Psi\rangle$, then the state vector of the system becomes coincident with (or is reduced to) the normalized projection $|\Psi\rangle$ onto the subspace spanned by the eigenvectors of **A** belonging to a_i. A literal interpretation of the prediction and reduction postulates together strongly suggests that a system in the state $|\Psi\rangle$ cannot meaningfully be said to "have a value" for A unless $|\Psi\rangle$ happens to coincide with one of the eigenvectors of **A**; indeed, this is the orthodox Copenhagen view.

Although there is a wide diversity of attitudes toward the Copenhagen interpretation, it is generally agreed that the prediction and reduction postulates together at least provide a correct algorithm for calculating, in a probabilistic sense, the results of virtually any conceivable series of quantum measurements. However, many physicists find the reduction postulate unpalatable, and they reject in particular the classically bizarre contention that a real physical system often does not have values for some of its legitimate observable variables. Some dissenters to Copenhagenism maintain that, leaving aside the reduction postulate, the implications of the prediction postulate can be fully accounted for by taking the following more plausible point of view[2]: The state vector $|\Psi\rangle$ specifies, not an individual system, but rather a statistical ensemble of identically prepared systems, each of which has precise values for *all* its dynamical observables. The values of these observables are distributed among the ensemble systems in such a way that the probability of randomly selecting an ensemble system that has $A = a_i$ is equal to the probability given by the prediction postulate for obtaining a_i in a measurement of A. We shall call this viewpoint a "simple ensemble interpretation" of quantum mechanics, and we note that it does *not* conflict with the Heisenberg uncertainty principle because it regards the uncertainty ΔA in A in the state $|\Psi\rangle$ as simply the statistical spread in the distribution of the **A** eigenvalues among the ensemble members.

The purpose of this paper is to show, in a way that should be accessible to first year students of quantum mechanics, that a simple ensemble interpretation is not possible for some quantum systems, and hence is untenable as a general premise. Our argument is patterned after Bell's[3] argument against simple hidden variable theories, or more precisely, Bell's argument as simplified and clarified by Wigner,[4] d'Espagnat,[5] and Harrison.[6] In a sense, we shall merely make the point that the difficulties discovered by Bell and his interpreters regarding hidden variable theories also arise for ensemble interpretations, despite the fact that ensemble interpretations need not overtly entail hidden variables.

To make our discussion as simple as possible, we shall

