What quantum measurements measure

Robert B. Griffiths*

Department of Physics, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213, USA

(Received 9 May 2017; published 13 September 2017)

A solution to the second measurement problem, determining what prior microscopic properties can be inferred from measurement outcomes ("pointer positions"), is worked out for projective and generalized (POVM) measurements, using consistent histories. The result supports the idea that equipment properly designed and calibrated reveals the properties it was designed to measure. Applications include Einstein's hemisphere and Wheeler's delayed choice paradoxes, and a method for analyzing weak measurements without recourse to weak values. Quantum measurements are noncontextual in the original sense employed by Bell and Mermin: if $[A,B] = [A,C] = 0$, $[B,C] \neq 0$, the outcome of an A measurement does not depend on whether it is measured with B or with C. An application to Bohm's model of the Einstein-Podolsky-Rosen situation suggests that a faulty understanding of quantum measurements is at the root of this paradox.

DOI: 10.1103/PhysRevA.96.032110

I. INTRODUCTION

A. The second measurement problem

The *measurement problem* is a central issue in quantum foundations, because textbook quantum mechanics uses the idea of a measurement to give a physical interpretation to probabilities generated from a quantum wave function, but never explains the measurement process itself in terms of more fundamental quantum principles. If, as is widely believed, quantum mechanics applies to macroscopic as well as microscopic phenomena, then it should be possible, at least in principle, to describe actual laboratory measurements in terms of basic quantum properties and processes, rather than employing "measurement" as an unanalyzed primitive.

It is convenient to divide the measurement problem into two parts. The *first measurement problem*, which is at the center of most discussions in the literature, is to understand how the measurement process can result in a well-defined macroscopic outcome or *pointer position*, to use the archaic but picturesque language of the foundations community, rather than some strange quantum superposition of the pointer in different positions, as results in many cases from a straightforward application of unitary time development: Schrödinger's equation leads to Schrödinger's cat. But even if the first measurement problem is solved, so that the pointer comes to rest at a single position, the *second measurement problem* remains: what can one infer from the pointer position regarding the microscopic situation that existed before the measurement took place, which the apparatus was designed to measure? Experimental physicists talk all the time about gamma rays triggering a detector, neutrinos arriving from the sun, and other microscopic objects or events which are invisible, and whose existence can only be inferred from the macroscopic outcomes of suitable measurements. Should we take this talk seriously? Maybe we do, but why, if the second measurement problem remains unresolved? Would we have any confidence in the stories told us by cosmologists if they did not understand the operation of their telescopes well enough to interpret the data these instruments provide?

A recent (and at the time of writing continuing) controversy [1,2] about the path followed by a photon passing through an interferometer on its way to a detector shows how difficult it is to analyze, using the tools of textbook quantum theory, with perhaps some additional *ad hoc* principles, a microscopic situation that is really not very complicated. This problem is, in turn, related to a hotly contested claim, published in a reputable journal, that information can be sent between two parties by means of a photon that is actually never—or at least hardly ever—present in the optical fiber that connects them [3–5]. What this suggests is that the failure of quantum physicists to solve the measurement problem(s) is not only an intellectual embarrassment—surely it is that, as pointed out by some leading physicists (see [6] and Sec. 3.7 of [7])—but also a serious impediment to ongoing research in areas such as quantum information, where understanding microscopic quantum properties and how they depend on time is central to the enterprise. In addition, a fuzzy understanding of quantum principles makes the subject hard to teach as well as to learn. Students confused by unfamiliar mathematics are not helped by the absence of a clear physical interpretation of what the mathematics means, something which neither textbooks nor instructors seem able to provide.

In this paper the second (and, incidentally, the first) measurement problem is addressed using the consistent histories, also known as decoherent histories, interpretation of quantum mechanics. While this approach is controversial (as is everything else in quantum foundations) it possesses specific principles and clear rules for applying and interpreting quantum theory at the microscopic level. These principles are comparatively few in number, include no reference to measurements, and apply universally to all quantum processes, whether microscopic or macroscopic, "from the quarks to the quasars." They are, so far as is known at present, consistent in the sense that when properly applied they do not lead to contradictions, and they have resolved (perhaps "tamed" would be a better term) various quantum paradoxes; see Chaps. 21 to 25 of [8] for a number of examples.

B. Article overview

The remainder of the paper is structured as follows. Section II explores the second measurement problem from a

*rgrif@cmu.edu

phenomenological perspective using two paradoxes, the first by Einstein and the second by Wheeler, that show why the problem is both difficult and confusing. Section II C is a brief discussion of how a measurement apparatus can be calibrated to ensure its reliability. Next a brief summary of the consistent histories approach, along with references to literature that provides further details, constitutes Sec. III; readers already familiar with consistent histories ideas can skip it.

Section IV is the heart of the paper, and contains the key ideas needed to address the second measurement problem both for projective measurements, Sec. IV A, and for generalized measurements (positive operator-valued measures, or POVMs), Sec. IV B. The emphasis is on simple cases of single measurements; situations where there are several successive measurement on the same system are not discussed, though the histories methodology can also be extended to such situations. A useful conceptual tool, which so far as we know has not been pointed out previously, is the backwards map from output (pointer) states to earlier microscopic properties. It is very helpful in identifying the microscopic properties which have been measured in the case of a generalized measurement. A separate Sec. IV C discusses nondestructive measurements and preparations, both closely related to von Neumann's measurement model. This may assist the reader in connecting the approach followed in this paper to ideas, such as wave function collapse, frequently encountered in textbook treatments and quantum foundations literature. The final section, Sec. IV D, has a few comments about density operators.

Next in Sec. V the tools developed in Sec. IV are applied to six different situations, where the first two, Secs. V A and V B, are closely related to the examples discussed earlier in Sec. II. The third, Sec. V C, is an elementary but not entirely trivial example of a POVM that is not a projective measurement. A fairly elementary, but again nontrivial, example in Sec. V D shows how a weak measurement can be interpreted in terms of quantum properties instead of the widely used "weak values." The last two applications address topics which often come up in the quantum foundations literature, and are hence somewhat controversial. It is argued in Sec. V E, using a less formal and more physical approach than [9], that if one uses Bell's original definition of "contextual," quantum mechanics is in fact noncontextual, despite confusing claims to the contrary. Finally, the Bohm (spin singlet) model of the famous Einstein-Podolsky-Rosen paradox is discussed in Sec. V F from the perspective of what one can infer from measurements on one of the spin-half particles about its prior properties and those of the other spin-half particle.

The final section, Sec. VI, is an attempt to summarize the most important conclusions about what it is that quantum measurements measure, while summarizing the principles which make it possible for the consistent histories interpretation to arrive at a satisfactory resolution of the second (as well as the first) measurement problem.

II. MEASUREMENT PHENOMENOLOGY

A. Einstein's paradox

Figure 1(a) shows Einstein's paradox (pp. 115–117 in Ref. [10], pp. 440–442 in Ref. [11]). A particle emerges from a small hole at the left and propagates as a spherical wave towards a curved fluorescent screen where its arrival is signaled by a flash of light at a particular point on the screen, a point which varies randomly on successive repetitions of the experiment. It seems as if the quantum wave collapses instantly when the particle reaches the screen, a result which bothered Einstein as it would mean a superluminal effect if every point on the screen is equidistant from the hole. An experimental physicist, on the other hand, might say that the particle travels on a straight line from the source to the screen, and could support that explanation by placing a collimator, a thick plate with a hole in it, between the source and the screen, and noting that now flashes are detected only at places on the screen which are connected to the source by a straight line passing through the hole, Fig. 1(b).

But isn't this second perspective classical, not quantum mechanical? No, for there is a good quantum mechanical description in which the particle is a small wave packet traveling from the source to the screen, Fig. 1(c); one only has to assume that the particle emerging from the source is described by such a wave packet whose initial direction of propagation is random from one run to the next. (And this gets around another problem with wave function collapse. If the particle reaches the screen, does this mean that its failing to interact with the collimator has collapsed the spherical wave enough so that it can fit through the hole?)

Continuing on, if the collimator has two holes, Fig. 2(a), one will observe flashes on the screen due to particles which have passed through one hole or the other, but never simultaneous

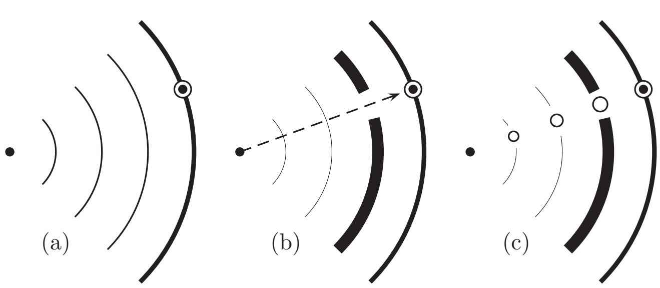

FIG. 1. Einstein paradox. (a) Spherical wave. (b) Particle moving on straight line through collimator. (c) Quantum wave packet passing through collimator.

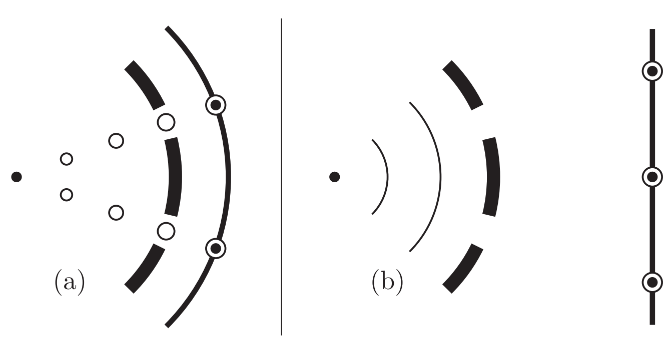

FIG. 2. (a) Collimator with two holes. (b) Fluorescent screen a large distance to the right of the collimator. Due to constructive interference of waves coming from the two holes a particle can sometimes be observed in a region which is classically forbidden.

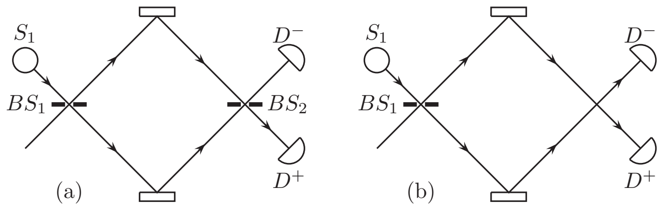

FIG. 3. Mach-Zehnder interferometer (a) with a source S_1, two beam splitters BS_1 and BS_2, and detectors D^+ and D^-; (b) with the second beam splitter removed.

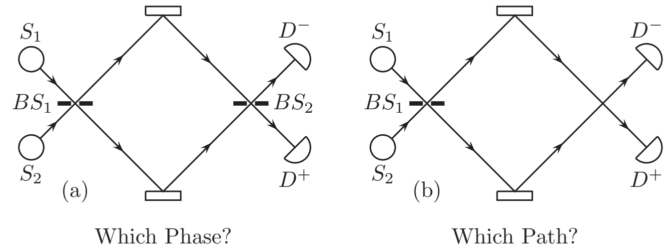

FIG. 4. Mach-Zehnder interferometer with two inputs (a) arranged to determine relative phase between the two arms; (b) arranged to measure which path (which arm).

flashes behind both holes. Again, easy to understand using the picture of little wave packets. But consider the situation in Fig. 2(b) where, if the two holes are formed very carefully and the fluorescent screen placed a long distance away, the result will be an interference pattern with the distance between fringes determined by, among other things, the distance between the two holes and the de Broglie wavelength of the quantum particle. The particle must, in this case, be thought of as a wave passing simultaneously through both holes and emerging behind them with a well-defined phase. We have arrived at the double-slit or two-hole paradox so well described by Feynman [12].

Everyone knows that quantum particles are waves, and quantum waves are particles. The gedanken experiments just discussed, especially the contrast between Figs. 2(a) and 2(b), illustrate the fact that sometimes a particle (fairly well localized wave packet) and sometimes a wave (coherence in phase over a macroscopic distance) description is needed in order to understand what is going on. The need to use different, and seemingly incompatible, descriptions is one of the fundamental difficulties behind the second measuring problem. One aim of the present article is to show how it can be addressed without invoking *retrocausation*: a future measurement influencing past behavior.

B. Mach-Zehnder with removable beam splitter

Einstein's paradox becomes easier to analyze if we consider the case of a Mach-Zehnder interferometer, Fig. 3(a), with an upper and lower arm connecting two beam splitters BS_1 and BS_2, and the phases adjusted so that a photon—hereafter referred to as a "particle"—from the source S_1 on the left is always detected by the lower detector D^+ on the right. That the particle is, in some sense at least, in both the upper and the lower arm while inside the interferometer can be checked by inserting two phase shifters, one in each arm. One then finds that, depending on the choice of phases, the particle will sometimes be detected in D^+ and sometimes in D^-. However, if both phases are identical, the particle will always be detected in D^+. Additional checks can be made by blocking either the upper arm or the lower arm, and noting that when one arm is blocked the particle will sometimes arrive in D^+ and sometimes in D^-.

If, on the other hand, the second beam splitter is absent, Fig. 3(b), the experimentalist will say that a particle detected in D^+ was originally in the upper arm of the interferometer, and if detected in D^- it was in the lower arm, as these are the direct paths from the first beam splitter to the detectors. This can be checked by placing barriers in the upper or lower arms of the interferometer and noting that a barrier in the upper arm will prevent the particle arriving at D^+, and one in the lower arm suppresses counts in D^-. Similarly, if a nondestructive measuring device, something which will register the particle's presence without seriously perturbing its motion, is placed in one of the arms, its outcome will show the expected correlation with the final detectors.

Wheeler's delayed choice paradox [13] comes from asking what will occur if just before the particle arrives at BS_2, when it has already passed BS_1 and is inside the interferometer, the second beam splitter is removed. Alternatively, suppose that the second beam splitter is absent while the particle is traversing the interferometer, but is suddenly inserted just before the particle arrives at the crossing point. One can imagine either of these experiments repeated many times, and the result will be that the presence or absence of BS_2 at the crossing point at the instant the particle arrives there determines whether the particle is always detected in D^+ or randomly detected in D^+ and D^-. And experimental checks can be carried out with phase shifters or barriers placed on the paths inside the interferometer. The paradox is perhaps most telling if one starts off with a series in which BS_2 is absent, and the particle arrives randomly in D^+ or D^-, so about half the time it is detected in D^-, and hence, plausibly, it has been following the lower path through the interferometer. Now undertake a series of runs in which BS_2 is initially absent, but is inserted in its proper place at the very last moment. In all of these runs the particle is detected by D^+. But in roughly half of these cases, assuming there is no retrocausal effect from the later insertion of BS_2, the particle must have been traveling through the lower arm, and were it traveling through the lower arm it would, upon passing through BS_2, arrive with equal probability in either of the detectors. Thus it might seem that sometimes the particle when traveling through the lower arm of the interferometer senses that at a future moment BS_2 will be present and decides to split itself into a pair of wave packets, one in each arm, with an appropriate phase, so that it will arrive with certainty at D^+. That seems very strange. Is there not some other way of understanding what is going on without invoking magic or retrocausation?

Adding a second source S_2 to Wheeler's paradox, Fig. 4, makes it somewhat analogous to our previous discussion of Einstein's paradox. In any given run, only one source emits a photon, and the phases have been chosen so that with the second beam splitter present a particle (photon) which originates in source S_1 will later arrive in D^+, and one emitted by S_2 will arrive at D^-. In both cases the particle while inside the interferometer is a superposition of a state $|z+\rangle$ in the upper arm and a state $|z-\rangle$ in the lower arm; in particular let us assume the phases are such that

$$S_1 \to (|z+\rangle + |z-\rangle)/\sqrt{2} \to D^+,$$
$$S_2 \to (|z+\rangle - |z-\rangle)/\sqrt{2} \to D^-. \quad (1)$$

One can then regard the second beam splitter and the two detectors as forming a single measurement apparatus that measures "which phase?"—the difference between the two possible relative phases, + vs − in Eq. (1)—when BS_2 is in place; or "which arm?" if BS_2 has been removed. Note the analogy with the situation depicted in Fig. 2 [with (a) and (b) interchanged]. The fact that in any particular run the experimenter, by leaving BS_2 in place or removing it, can measure which phase or which path but cannot determine both, is a fundamental fact of quantum mechanics. Taking it into account is essential if one is to make progress in resolving the second measurement problem.

C. Calibration

Competent experimenters check their apparatus in various ways to make sure it is operating as designed and gives reliable results. There are varieties of tests, some suggested earlier: placing collimators in various places, removing beam splitters from a Mach-Zehnder interferometer, placing absorbers in its arms, etc. If the apparatus is designed to measure the value of some quantity (observable) A associated with a particle, the simplest form of *calibration* means preparing many particles with known values of A, thus having the property corresponding to some particular eigenvalue, and seeing whether the measurement outcome (pointer position) corresponds in each case to the known property. Once the calibration has been carried out the experimenter can be confident that when a particle of this type is measured by the apparatus, the outcome will indicate the value of A possessed by the particle just before it reached the apparatus, even when the particle's prior history is unknown. Experimenters frequently make assumptions of this kind, and without it a significant part of experimental physics would be impossible. A proper quantum mechanical theory of measurement must be able to justify this practice. In reality things are not always so simple, since the apparatus is never perfect and one may have to account for possible errors; however, for the present discussion we shall focus on the ideal case in order to get to the essentials of quantum measurements.

III. PROPERTIES, PROBABILITIES, AND HISTORIES

This section contains a rapid review of material found elsewhere; readers familiar with consistent histories can skip ahead to Sec. IV. See [14] for an introduction to consistent histories, [8] for a detailed treatment, and [15] for extended comments on some conceptual difficulties.

A. Quantum properties

We use the term *physical property* for something like "the energy is less than 2 joules" or "the particle is in a region R in space," something which can be true or false, and thus distinct from a *physical variable* such as the energy or position, represented by a real number in suitable units. Von Neumann, Sec. III.5 of [16], proposed that a quantum property should correspond to a *subspace* of the quantum Hilbert space, or, equivalently, the *projector* (orthogonal projection operator) onto this subspace. (We are only concerned here with finite-dimensional Hilbert spaces for which all subspaces are closed.) What one finds in textbooks is consistent with von Neumann's prescription, though this is not always clearly stated.

A projector, a Hermitian operator equal to its square, is the quantum analog of an *indicator function* $P(\gamma)$ on a classical phase space Γ, a function that takes the value 1 if at the point γ the corresponding physical property is true, or 0 if it is false. For example, the property that the energy of a harmonic oscillator is less than 2 joules corresponds to an indicator $P(\gamma)$ equal to 1 for γ inside, and 0 for γ outside, an ellipse centered at the origin of the (x,p) phase plane. A quantum projector's eigenvalues are 1 or 0, which supports the analogy with a classical indicator. One can make a plausible case that any "classical" property of a macroscopic physical object, when viewed in quantum terms, is represented by a quantum projector on a very high-dimensional subspace of an enormous Hilbert space.

The smallest nontrivial quantum subspace is one-dimensional, consisting of all complex multiples of a normalized ket $|\psi\rangle$, and the projector is given by the corresponding Dirac dyad

$$[\psi] = |\psi\rangle\langle\psi|. \quad (2)$$

We will often make use of this convenient square bracket notation. A projector on a two-dimensional subspace can be written in the form $[\psi^0] + [\psi^1]$, where $|\psi^0\rangle$ and $|\psi^1\rangle$ form an orthonormal basis for the subspace, and similarly for larger subspaces.

The analogy between quantum projectors and classical indicators also works for *negation*. The projector corresponding to the property "NOT P" is $I - P$, where I is the identity operator, and the same holds for a classical indicator when I is understood as the function taking the value 1 everywhere on the phase space. Given two indicator functions representing properties P and Q, their product, which is obviously the same written in either order, $P(\gamma)Q(\gamma) = Q(\gamma)P(\gamma)$, is the indicator for the property P AND Q. (Think of "energy less than one joule" AND "momentum is positive"). But in the quantum world the product of two projectors P and Q is itself a projector *if and only if* they commute: $PQ = QP$, and in this case the product can be associated with the property P AND Q.

But suppose that P and Q do *not* commute; what then? Consider a specific example, that of a spin-half particle, where the Hilbert space is two-dimensional, and spanned by two orthonormal kets $|z^+\rangle$ and $|z^-\rangle$, eigenvectors of S_z, the z component of spin angular momentum, with eigenvalues $+1/2$ and $-1/2$ in units of $\hbar$. The projectors

$$P^+ = [z^+], \quad P^- = [z^-], \quad (3)$$

in the notation used in Eq. (2), represent these two physical properties; they commute and their product is 0. Similarly,

$$|x^+\rangle = (\,|z^+\rangle + |z^-\rangle\,)/\sqrt{2}, \quad |x^-\rangle = (\,|z^+\rangle - |z^-\rangle\,)/\sqrt{2} \tag{4}$$

are eigenvectors corresponding to the eigenvalues $+1/2$ and $-1/2$ of the x component of spin angular momentum S_x. The corresponding projectors

$$Q^+ = [x^+], \quad Q^- = [x^-] \tag{5}$$

commute, and their product is zero. However, neither Q^+ nor Q^- commutes with either P^+ or P^-. Because the projectors do not commute there is, in the consistent histories approach, no way to make sense of a statement like "$S_z = +1/2$ AND $S_x = -1/2$." And there is no nontrivial subspace of the Hilbert space which can be associated with such a combination. (In quantum logic [17,18] one would associate the trivial subspace containing only the 0 ket with such a conjunction, but quantum logic has its own set of conceptual difficulties; see [15].) This is an instance of the *single-framework rule* discussed in more detail in Sec. III C.

From time to time the claim has been made that the consistent histories approach is logically inconsistent. However, none of these claims when scrutinized has turned out to be correct. What typically happens is that the author has either overlooked the single-framework rule or has not taken it seriously. Arguments that show that consistent histories is internally consistent will be found in Chap. 16 of [8], Sec. 4.1 of [15], and Sec. 8.1 of [19].

B. Quantum probabilities

Ordinary (Kolmogorov) probability theory employs a *sample space* of *mutually exclusive* items or situations which together *exhaust all possibilities*, and an *event algebra* which in simple situations consists of all subsets (including the empty set) of items from the sample space. In classical statistical mechanics the sample space can consist of all the distinct points γ that make up the phase space Γ, but one could also cut up the phase space into nonoverlapping regions, "cells," and use these for the sample space. The quantum analog of a sample space is a *projective decomposition of the identity* (PDI): a collection of projectors $\{P^j\}$ (the superscripts are labels, not exponents) satisfying

$$I = \sum_j P^j, \quad P^j = (P^j)^\dagger, \quad P^j P^k = \delta_{jk} P^j. \tag{6}$$

Obviously, each projector commutes with every other projector in the PDI. The simplest choice for a corresponding event algebra, one which will suffice for our purposes, consists of the 0 projector, all projectors belonging to the PDI, and in addition all sums of two or more distinct projectors from the PDI.

Given a physical variable A represented by a Hermitian operator A (there is no harm in using the same symbol for both) there is an associated PDI employed for the spectral decomposition of A,

$$A = \sum_j \alpha_j P^j, \tag{7}$$

where the eigenvalues α_j are the possible values which A can take on, and P^j identifies the subspace where A takes on the value α_j. [We assume that $\alpha_j \neq \alpha_k$ if $j \neq k$ in Eq. (7); thus for degenerate eigenvalues the corresponding P^j may project onto a space of dimension greater than 1.]

In classical physics it is usually the case that only a single sample space need be considered when discussing a particular physical problem, and so its choice needs no emphasis, and it may not even be mentioned. In quantum physics this is no longer the case: many mistakes and numerous paradoxes, e.g., the Kochen-Specker paradox (see Sec. V E), are based on not paying sufficient attention to the sample space in circumstances in which several distinct and incompatible sample spaces may seem like reasonable choices. For this reason it is convenient to use a special term, *framework*, to indicate the sample space or the corresponding event algebra which is under discussion.

A central feature of consistent histories is the *single-framework rule*, which states that probabilistic reasoning in the quantum context must always be carried out using a specific and well-defined framework. This rule does *not* prevent the physicist from *using* many different frameworks when analyzing a particular physical problem; instead it prohibits *combining* results from *incompatible* frameworks. Two PDIs $\{P^j\}$ and $\{Q^k\}$ and the corresponding event algebras are *compatible* provided that all the projectors in one commute with all the projectors in the other: $P^jQ^k = Q^kP^j$ for every j and k. In this case there is a *common refinement*, a PDI consisting of all nonzero products of the form P^jQ^k. Otherwise the frameworks are incompatible, and the single-framework rule prohibits combining a (probabilistic) inference made using one framework with another that employs a different framework. If the two frameworks are compatible, then inferences in one can be combined with those in the other using the common refinement, which contains both of the event algebras, so again only a single framework is required. (An additional requirement—consistency conditions—for combining frameworks arises in the case of quantum histories; Sec. III C.)

A PDI can be assigned a probability distribution $p_j = \Pr(P^j)$, where the p_j are nonnegative real numbers that sum to 1, and this distribution will generate the probabilities for all the elements in the corresponding event algebra, just as in ordinary probability theory; e.g., the property $P^1 + P^3$ is assigned the probability $p_1 + p_3$. In quantum mechanics there are various schemes for assigning probabilities. One method starts with a wave function or pure quantum state $|\psi\rangle$, and assigns to the elements of a PDI $\{P^j\}$ probabilities

$$p_j = \langle\psi|P^j|\psi\rangle = \mathrm{Tr}([\psi]P^j). \tag{8}$$

In this situation it is helpful to refer to $|\psi\rangle$ as a *pre-probability*; i.e., it is used to construct a probability distribution. Since probability distributions are generally not considered part of physical reality, at least not in the same sense as physical properties, a ket or wave function used in this way need not be interpreted as something physical; instead it is simply a tool used to compute probabilities. But in some other context $|\psi\rangle$ may be a way of referring to the property represented by the projector $[\psi]$. Carelessly combining these two usages can cause a great deal of confusion. Note in particular that as long as two of the p_j in Eq. (8) are nonzero, the property

$[\psi]$, or to be more precise the minimal PDI $\{[\psi], I-[\psi]\}$ that contains it, is incompatible with the PDI $\{P^j\}$. Hence the single-framework rule prevents using $|\psi\rangle$ as a pre-probability, as in Eq. (8), while at the same time regarding it as a physical property of the quantum system.

Since the consistent histories interpretation of quantum theory allows many distinct but incompatible frameworks, a natural question is, Which is the *right* framework to use in describing some situation of physical interest? In thinking about this it is helpful to remember that a fundamental difference between classical and quantum mechanics is that the former employs a phase space and the latter a Hilbert space for describing a physical system. At a single time a single point in the phase space represents the "actual" state of a classical system: all properties (subsets of points in the phase space) which contain this point are *true* and all which do not contain the point are *false*. The term *unicity* has been used in Sec. 27.3 of [8] and in Refs. [14,15] to describe this concept of a single unique state of affairs at any given time. However, in the quantum Hilbert space the closest analogy to a single point in classical phase space is a one-dimensional subspace or *ray*. If one assumes that one particular ray is true, then one might suppose that all rays orthogonal to it are false. But there are many rays that are neither identical to nor orthogonal to the ray in question; what shall be said of them? Thus attempting to extend the concept of unicity into the quantum domain runs into problems. We have good reason to believe that physical reality is better described by quantum theory than by classical physics, and hence certain classical concepts must be abandoned, to join others, such as the earth immobile at the center of the universe, which modern science has rendered untenable, even though for certain purposes they may remain useful approximations. Unicity seems to belong to that category.

But the question remains: what are the criteria which lead to the use of a particular framework, rather than another which is incompatible with it? The examples in Sec. II and various applications in Sec. V suggest that quantum physical situations possess what one might call *different aspects*, and a quantum description of a particular aspect can only be constructed using a framework compatible with that aspect. For example, the S_z "aspect" of a spin-half particle can only be discussed using the S_z framework; the S_x framework is of no use. As is usual with with unfamiliar concepts, the best way to understand them is to apply them to several different examples. In particular, in Secs. V A and V B we will show how the use of frameworks can "untangle" the paradoxes in Secs. II A and II B.

C. Histories and the extended Born rule

A quantum history is best understood as a *sequence of quantum properties at successive times*. A classical analogy is a sequence of coin tosses, or rolls of dice. The theory is simplest if one employs a finite set of discrete times, rather than continuous time. This is no real limitation, as these times may be arbitrarily close together. A history associated with the times $t_0 < t_1 < t_2 < \cdots t_n$ can be written in the form

$$Y = F_0 \odot F_1 \odot F_2 \odot \cdots F_n, \tag{9}$$

where each F_j is a projector representing some quantum property at the time t_j, and the $\odot$ separating properties at successive times are tensor product symbols, a variant of $\otimes$. Thus if $\mathcal{H}$ is the quantum Hilbert space at one time, Y in Eq. (9) is a projector on the tensor product *history* Hilbert space $\breve{\mathcal{H}} = \mathcal{H}^{\otimes(n+1)}$. A *family* of histories is a collection of such projectors that sum to the history identity $\breve{I} = I \odot I \odot \cdots I$, thus a PDI. For present purposes it suffices to use a family in which the histories are of the form

$$Y^\alpha = [\Psi_0] \odot F_1^{\alpha_1} \odot F_2^{\alpha_2} \odot \cdots F_N^{\alpha_n}, \tag{10}$$

where $[\Psi_0]$, see (2), is the projector on a pure state $|\Psi_0\rangle$. The superscripts are labels distinguishing different projectors at the same time, and together they form a vector $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$. In addition there is a special history $Y^0 = I - [\Psi_0] \odot I \odot I \cdots I$ which is assigned zero probability, and whose sole purpose is to ensure that the history projectors sum to $\breve{I}$.

A *complete* family of histories is one in which the Y^α sum to $\breve{I}$, but we will also use the term if they sum to $\breve{I} - Y^0$. One way to ensure that the family is complete is if for each time $t_j > t_0$ it is the case that the $\{F_j^{\alpha_j}\}$ are a PDI of $\mathcal{H}$, but this is often too restrictive. There is no reason why a family should not contain projectors on states "entangled" between different times, but in the following discussion we will only need "product" histories as in Eq. (9).

Since a family of histories is a PDI it can serve as a probabilistic sample space for the quantum analog of a classical stochastic process such as a random walk. As in the classical case there is no fixed rule for assigning probabilities to such a process. However, in a *closed* quantum system for which Schrödinger's equation yields unitary time development operators $T(t',t)$ (e.g., $\exp[-i(t'-t)H/\hbar]$ in the case of a time-independent Hamiltonian H) these can be used to assign probabilities to a history family using an extension of the Born rule, provided certain *consistency* (or *decoherence*) *conditions* are satisfied. If all histories start with the same initial pure state one defines a *chain ket* (an element of $\mathcal{H}$ not $\breve{\mathcal{H}}$):

$$|Y^\alpha\rangle = F_n^{\alpha_n} T(t_n,t_{n-1}) F_{n-1}^{\alpha_{n-1}} T(t_{n-1},t_{n-2}) \cdots F_1^{\alpha_1} T(t_1,t_0)|\Psi_0\rangle. \tag{11}$$

The consistency conditions are the requirement that the chain kets are orthogonal for distinct histories,

$$\langle Y^\alpha | Y^{\alpha'}\rangle = 0 \quad \text{for } \alpha \neq \alpha'. \tag{12}$$

When it is satisfied the extended Born rule assigns to each history of the sample space a probability

$$\Pr(Y^\alpha) = \langle Y^\alpha | Y^\alpha\rangle. \tag{13}$$

The orthogonality requirement (12) is not unnatural when one remembers that the $|Y^\alpha\rangle$ are elements of the single-time Hilbert space $\mathcal{H}$, not the history space $\breve{\mathcal{H}}$, and the ordinary Born rule is used to assign probabilities to an orthonormal basis, or, more generally, a PDI. In fact, for a history involving only two times, t_0 and t_1, the consistency condition is automatically satisfied because the $F_1^{\alpha_1}$ for different α_1 form a PDI on $\mathcal{H}$, and then (13) is just the usual Born probability.

It is important to notice that quantum mechanics allows a *description* of what happens in an individual realization of a quantum stochastic process, even though the dynamics is probabilistic, the same as in a classical stochastic theory. One is sometimes given the impression that quantum theory only allows a discussion of statistical averages over many runs of an experiment. This is not the case, and it is easy to identify instances where individual outcomes and not just averages play a significant role. For example, if Shor's quantum algorithm [20,21] is employed to factor a long integer, then at the end of each run the outcome of a measurement is processed to see whether this result solves the problem, and if it does, no further runs are needed. While it may take more than one run to achieve success, the outcome of a particular run is a significant quantity, and not just the average over several runs. Similarly, in the case of Einstein's paradox, Sec. II A, a flash of light at a particular point on the fluorescent screen, Fig. 1(a), can be understood to mean that the particle traveled on a straight (or almost straight) path from the source to the screen on this particular occasion.

IV. MEASUREMENT MODELS

A. Projective measurements

Our first model is a generalization of the one introduced by von Neumann in Sec. VI.3 of [16].[1] Let $\mathcal{H}_s$ be the Hilbert space of the system to be measured, which for convenience will hereafter be referred to as "the particle," whereas the measuring apparatus, including its environment if that is significant, is described by a Hilbert space $\mathcal{H}_m$. The total system with Hilbert space $\mathcal{H}_M = \mathcal{H}_s \otimes \mathcal{H}_m$ is thought of as closed, so its dynamics can be associated with a collection of unitary time development operators $T(t',t)$. We will focus on histories involving three times $t_0 < t_1 < t_2$, where the interval from t_0 to t_1 is so short that $T(t_1,t_0) \approx I$ and thus

$$T(t_2,t_0) \approx T(t_2,t_1) \tag{14}$$

with negligible error. At the initial time t_0 the particle can be assigned a quantum state $|\psi_0\rangle$ in $\mathcal{H}_s$, and the apparatus (and environment) a state $|\Omega_0\rangle$ in $\mathcal{H}_m$; hence an initial state

$$|\Psi_0\rangle = |\psi_0\rangle \otimes |\Omega_0\rangle \tag{15}$$

for the combined, closed system. The use of pure states rather than density operators does not involve any loss of generality; see Sec. IV D for additional comments. But the requirement that $|\Psi_0\rangle$ in Eq. (15) be a *product state* is important. It means that the particle and the apparatus (or environment) are initially uncorrelated, at least to a sufficiently good approximation.

We assume that the interaction between the particle and the apparatus takes place during the time interval between t_1 and t_2, and as a consequence of this interaction

$$T(t_2,t_1)(|s^j\rangle \otimes |\Omega_0\rangle) = |\Phi^j\rangle, \tag{16}$$

where the $|s^j\rangle$ form an orthonormal basis for the particle Hilbert space $\mathcal{H}_s$, while the $|\Phi^j\rangle$, which lie in the Hilbert space $\mathcal{H}_M$, are states of the particle plus apparatus that correspond to distinct macroscopic outcomes of the measurement—distinct "pointer positions" of the apparatus, to use the traditional terminology of quantum foundations—in the sense of satisfying (17) below. (The space $\mathcal{H}_M$ has the same dimension as $\mathcal{H}_s \otimes \mathcal{H}_m$, but we have not written it in that form since sometimes the particle does not even exist at the end of the measurement. See the discussion of nondestructive measurements in Sec. IV C.) These pointer positions are mutually orthogonal, as is always the case for states which are macroscopically distinct. To be more precise, we assume there is a PDI $\{M^k\}$ on $\mathcal{H}_M$ such that

$$M^k|\Phi^j\rangle = \delta_{jk}|\Phi^j\rangle, \tag{17}$$

where each M^k is a projector on a macroscopic subspace (property) whose interpretation is that the pointer is in position k, and (17) says that $|\Phi^k\rangle$ lies within the subspace defined by M^k. To ensure that the $\{M^k\}$ sum to the identity on $\mathcal{H}_M$, assume that the possible pointer positions are represented by $k = 1,2,\ldots,n$, and let

$$M^0 := I_M - \sum_{k=1}^{n} M^k \tag{18}$$

project on the subspace that includes all other possibilities (e.g., the apparatus has broken down).

To better understand what this measurement measures it is useful to introduce an *isometry* $J : \mathcal{H}_s \to \mathcal{H}_M$ defined by

$$J|\psi\rangle = T(t_2,t_1)(|\psi\rangle \otimes |\Omega_0\rangle). \tag{19}$$

An isometry, like a unitary, preserves lengths, and is characterized by the requirement that

$$J^\dagger J = I_s, \tag{20}$$

where $J^\dagger : \mathcal{H}_M \to \mathcal{H}_s$ is the adjoint of J. (The operator $JJ^\dagger : \mathcal{H}_M \to \mathcal{H}_M$ is a projector on the subspace of $\mathcal{H}_M$ that is the image of under J of $\mathcal{H}_s$, and is not important for our discussion.)

The isometry that corresponds to $T(t_2,t_1)$ in Eq. (16) is

$$J|s^j\rangle = |\Phi^j\rangle. \tag{21}$$

Combining this with (17) leads to

$$M^k J|s^j\rangle = \delta_{jk} J|s^j\rangle. \tag{22}$$

Multiplying both sides on the left by $J^\dagger$ and using (20) yields

$$J^\dagger M^k J|s^j\rangle = \delta_{jk}|s^j\rangle, \tag{23}$$

which implies that

$$[s^k] = |s^k\rangle\langle s^k| = J^\dagger M^k J. \tag{24}$$

That is, the "backwards map" $J^\dagger(\cdot)J$ applied to the projector M^k on the subspace that corresponds to pointer position k is the prior microscopic state $[s^k]$ giving rise to this outcome.

To complete the discussion of projective measurements we need to introduce families of histories. Let us begin with the family $\{Y^k\}$ consisting of histories

$$Y^k = [\Psi_0] \odot I \odot M^k \tag{25}$$

at times $t_0 < t_1 < t_2$, where $|\Psi_0\rangle$ was defined in Eq. (15), and

$$|\psi_0\rangle = \sum_j c_j |s^j\rangle \tag{26}$$

[1] Von Neumann also gives a specific application of his general model to the case of a "Gaussian probe" whose momentum is shifted by an (almost) instantaneous interaction with the measured system. Our discussion concerns the more general model rather than its application to the Gaussian probe.

is an arbitrary state of $\mathcal{H}_s$. The chain kets

$$|Y^k\rangle = c_k|\Phi^k\rangle \tag{27}$$

associated with these histories [remember that $T(t_2,t_1) \approx T(t_2,t_0)$] are obviously orthogonal to each other in view of (17) and the fact that the $\{M^k\}$ form a PDI. Thus the Born rule assigns a probability

$$\Pr\left(M_2^k\right) = \langle Y^k|Y^k\rangle = |c_k|^2, \tag{28}$$

the absolute square of the coefficient of $|\psi_0\rangle$ in Eq. (26), to the pointer outcome k, in agreement with textbooks, but without employing any special rule for measurements, since (28) is nothing but a particular application of the general formula (13) that assigns probabilities to histories.

Note that the *first* measurement problem, attempting to give a physical interpretation to the macroscopic superposition state

$$|\Psi_2\rangle = T(t_2,t_0)|\Psi_0\rangle = \sum_j c_j|\Phi^j\rangle, \tag{29}$$

never arises, because $|\Psi_2\rangle$ has never entered the discussion. To be sure, from the consistent histories perspective there is nothing wrong with the family consisting of just the two histories

$$[\Psi_0] \odot I \odot \{[\Psi_2], I - [\Psi_2]\}, \tag{30}$$

where each history uses one of the projectors inside the curly brackets. It (trivially) satisfies the consistency condition, and the Born rule assigns a probability of 1 to $[\Psi_2]$. It is a perfectly good quantum description which, however, is incompatible with the family (25) if at least two of the c_j in Eq. (26) are nonzero, since $[\Psi_2]$ will then not commute with the corresponding M^j, rendering a discussion of measurement outcomes impossible. Combining the families in Eqs. (25) and (30) is as silly as simultaneously assigning to a spin-half particle a value for S_z along with one for S_x. The choice of which of these families to use will generally be made on pragmatic grounds. In particular, if one wants to discuss real experiments of the sort actually carried out in laboratories and what one can infer from their outcomes—one might call this practical physics—the choice is clear: one needs to employ a family in which measurements have outcomes.

There are physicists who object to a framework choice based on pragmatic grounds which seem related to human choice, e.g., see Sec. 3.7 of [7], though they might not object to astronomers interested in the properties of Jupiter using concepts appropriate to that planet rather than, say, Mars. Of course this is a classical analogy, but thinking about it, along with the spin-half example mentioned earlier, may help in understanding how the single-framework rule can assist in sorting out quantum paradoxes while still allowing quantum theory to be an objective science. The idea that there can only be exactly one valid quantum description, the principle of unicity discussed in Sec. III B, runs into difficulties in the case of Einstein's paradox, Sec. II A, as well rendering the infamous first measurement problem insoluble for reasons that have just been discussed.

After this diversion let us return to the second measurement problem. To see how the macroscopic measurement outcomes M^k are related to the microscopic properties the measurement was designed to measure, we introduce a refinement $\{Y^{jk}\}$,

$$Y^{jk} = [\Psi_0] \odot [s^j] \odot M^k, \tag{31}$$

of the family (25) considered previously. Here $[s^j]$ at the intermediate time t_1 is to be interpreted, following the usual physicists' convention, as $[s^j] \otimes I_m$; the property $[s^j]$ of the particle and no information about anything else. The corresponding chain kets, see (26) and (16),

$$|Y^{jk}\rangle = c_j\delta_{jk}|\Phi^k\rangle, \tag{32}$$

are mutually orthogonal since the $|\Phi^k\rangle$ are orthogonal. Thus the family $\{Y^{jk}\}$ is consistent, and yields a joint probability distribution

$$\Pr\left(s_1^j, M_2^k\right) = \langle Y^{jk}|Y^{jk}\rangle = \delta_{jk}|c_j|^2, \tag{33}$$

where the subscripts of the arguments of Pr( ) indicate time. Summing over j gives (28), and combining that with (33) yields conditional probabilities:

$$\Pr\left(s_1^j \mid M_2^k\right) = \Pr\left(s_1^j, M_2^k\right) / \Pr\left(M_2^k\right) = \delta_{jk}, \tag{34}$$

assuming c_k is nonzero. In words: if the measurement outcome (pointer position) is k, i.e., M^k, at time t_2, the particle certainly had the property $[s^k]$ at time t_1. Thus the second measurement problem has been solved for the case of projective measurements. Note that this conclusion does *not* depend upon the initial state $|\psi_0\rangle$, which only determines the probability of the measurement outcome M^k as noted above in Eq. (28). [If $c_k = 0$, (34) does not hold, but it is also not needed, since the outcome k will never occur.]

B. Generalized measurements and POVMs

The basic setup for discussing generalized measurements is the same as that in Sec. IV A: times $t_0 < t_1 < t_2$, an initial state (15) at time t_0, negligible time development [see (14)] between t_0 and t_1, the isometry J defined in Eq. (19), and a PDI $\{M^k\}$ corresponding to different pointer positions at t_2. However, we now drop the assumption of an orthonormal basis $\{|s^j\rangle\}$ of $\mathcal{H}_s$ with $J|s^j\rangle$ lying in the space M^j. Instead, use the backwards map of the projectors on the pointer subspaces to *define* for each k an operator

$$Q^k := J^\dagger M^k J \tag{35}$$

on $\mathcal{H}_s$. For a projective measurement $Q^k = [s^k]$ is the property possessed by the particle at the earlier time t_1 when the measurement outcome is M^k, and we shall see that something similar, though a bit more complicated, holds for generalized measurements. Another special case, a *generalized projective measurement*, is one in which each Q^k is a projector and the $\{Q^k\}$ form a PDI, but one or more may have a rank (so project on a subspace of dimension) greater than 1.

The collection $\{Q^k\}$ forms a POVM (positive operator-valued measure), a collection of positive semidefinite operators with sum equal to the identity on $\mathcal{H}_s$. The equality

$$\langle\psi|Q^k|\psi\rangle = \langle\psi|J^\dagger M^k J|\psi\rangle = \langle\Psi|M^k|\Psi\rangle \geqslant 0, \tag{36}$$

for an arbitrary $|\psi\rangle$ in $\mathcal{H}_s$, with $|\Psi\rangle = J|\psi\rangle$, demonstrates that Q^k, just like the projector M^k, is a positive semidefinite operator. Summing both sides of (36) over k and remembering that the M^k form a PDI shows that

$$\sum_k Q^k = I_s, \tag{37}$$

completing the proof that $\{Q^k\}$ is a POVM. [Note that the special M^0 in Eq. (18) gives rise to $Q^0 = 0$.]

The first measurement problem for such a POVM is solved in exactly the same way as for the von Neumann model: use the PDI $\{M^k\}$ at time t_2, not the projector $[\Phi_2]$ of the unitarily evolved state. The second measurement problem is more subtle, as it requires introducing suitable properties as events at t_1 to produce a consistent family. The choice is not unique, but the following is a quite general and fairly useful approach. The spectral decomposition of Q^k can be written in the form

$$Q^k = \sum_j q_{jk}\xi^{jk}, \quad \sum_j \xi^{jk} = I_s, \tag{38}$$

where for each fixed k the ξ^{jk} labeled by j are projectors that form a PDI on $\mathcal{H}_s$, while the $q_{jk} \geqslant 0$ are the corresponding eigenvalues of Q^k. We assume the eigenvalues are unique, $q_{jk} \neq q_{j'k}$ when $j \neq j'$, so some of the ξ^{jk} may have rank greater than one. As with any PDI the projectors are orthogonal and sum to the identity:

$$\xi^{jk}\xi^{j'k} = \delta_{jj'}\xi^{jk}, \quad \sum_j \xi^{jk} = I_s. \tag{39}$$

The family $\{Y^{jk}\}$ of histories

$$Y^{jk} = [\Psi_0] \odot \xi^{jk} \odot M^k \tag{40}$$

when augmented with the uninteresting $[\Psi_0] \odot I \odot M^0$ (of zero weight) is complete, since

$$\sum_j Y^{jk} = [\Psi_0] \odot I \odot M^k. \tag{41}$$

The chain kets

$$|Y^{jk}\rangle = M^k J\, \xi^{jk}\, |\psi_0\rangle \tag{42}$$

are obviously mutually orthogonal if the two k values differ. For a given k we need to consider

$$\begin{aligned}\langle Y^{jk}|Y^{j'k}\rangle &= \langle\psi_0|\xi^{jk}J^\dagger M^k J\xi^{j'k}|\psi_0\rangle\\ &= \langle\psi_0|\xi^{jk}Q^k\xi^{j'k}|\psi_0\rangle\\ &= \delta_{jj'}q_{jk}\langle\psi_0|\xi^{jk}|\psi_0\rangle,\end{aligned} \tag{43}$$

where the second equality follows from (35), the third from (38) and (39). Thus the family $\{Y^{jk}\}$ defined in Eq. (40) is consistent, with probabilities

$$\Pr\left(\xi_1^{jk}, M_2^{k'}\right) = \delta_{kk'}\, q_{jk}\, \langle\psi_0|\xi^{jk}|\psi_0\rangle, \tag{44}$$

where subscripts 1 and 2 identify the times t_1 and t_2 before and after the measurement takes place. It follows that

$$\Pr\left(M_2^k\right) = \sum_j \Pr\left(\xi_1^{jk}, M_2^k\right) = \langle\psi_0|Q^k|\psi_0\rangle, \tag{45}$$

$$\Pr\left(\xi_1^{jk'}\,\middle|\,M_2^k\right) = \delta_{kk'}\, q_{jk}\, \langle\psi_0|\xi^{jk}|\psi_0\rangle / \langle\psi_0|Q^k|\psi_0\rangle. \tag{46}$$

What (46) tells us is that if the outcome (pointer position) is k the system earlier had one of the properties ξ^{jk}, with probabilities that will in general depend on the initial particle state $|\psi_0\rangle$. If Q^k is itself a projector or proportional to a projector, as will be the case for a general projective measurement, one can be sure that the particle possessed the property Q^k at time t_1. If the support of Q^k is a proper subspace of $\mathcal{H}_s$, the system can be assigned the property corresponding to this subspace at the time immediately before the measurement. If neither of these conditions holds it may be possible on the basis of additional information about $|\psi_0\rangle$ to assign probabilities to the different ξ^{jk} for this k, or perhaps argue that some of these probabilities are negligible, allowing one with reasonable confidence to say something nontrivial about the property possessed earlier by the particle.

Note that whereas for a fixed k the ξ^{jk} for different j are mutually orthogonal, for *different* k values, different outcomes of the experiment, one may be able to draw different and perhaps mutually incompatible conclusions about the prior properties. This is a feature of quantum measurements which has given rise to a lot of confusion, and is best discussed in terms of a specific example; see the one in Sec. V C. While the consistent family in Eq. (40) is not the only possibility for discussing what one can learn about the prior state of the particle from measurement outcomes, it is a rather natural choice, especially when nothing else is known about the measured system.

C. Nondestructive measurements and preparations

A *measurement* determines a past property whereas a *preparation* is a procedure to prepare a particular quantum state, and a *nondestructive* measurement combines the two: the apparatus both measures and prepares certain properties. While preparations lie somewhat outside the scope of the present paper, it is worthwhile making some remarks on the subject, if only because of the confusion found in textbooks and other publications, where "measurement" is often (incorrectly) defined as something that has to do with "wave function collapse." The confusion goes back to von Neumann's original measurement model in which, using the notation of the present paper, $\mathcal{H}_M = \mathcal{H}_s \otimes \mathcal{H}_m$, and the isometry J in Eq. (19) takes the form

$$J|s^j\rangle = |s^j\rangle \otimes |\Phi^j\rangle, \quad M^k|\Phi^j\rangle = \delta_{jk}|\Phi^j\rangle, \tag{47}$$

with the $\{|s^j\rangle\}$ an orthonormal basis of $\mathcal{H}_s$. (The $|\Phi^j\rangle$ and the PDI $\{M^k\}$ now refer to $\mathcal{H}_m$ rather than $\mathcal{H}_M$, as in our earlier discussion, but this is a minor difference.) In place of (31) use the family

$$Y^{jj'k} = [\psi_0] \otimes [\Omega_0] \odot \{[s^j]\} \odot \{[s^{j'}] \otimes [M^k]\}. \tag{48}$$

It is straightforward to show that it is consistent, since all the chain kets vanish except for the cases $j = j' = k$, with the result

$$\Pr\left(s_1^j, s_2^{j'}, M_2^k\right) = \delta_{jj'}\delta_{jk}\langle\psi_0|\,[s^j]\,|\psi_0\rangle, \tag{49}$$

$$\Pr\left(s_1^j\,\middle|\,M_2^k\right) = \delta_{jk}, \quad \Pr\left(s_2^j\,\middle|\,M_2^k\right) = \delta_{jk}. \tag{50}$$

This measurement is nondestructive in the sense that from the outcome M^k one can immediately infer that the particle property both before and after the measurement was $[s^k]$, so it did not change. Furthermore, this conclusion is independent of the initial particle state $|\psi_0\rangle$ [assuming only that c_k in Eq. (26) is not zero; if it is zero the outcome M^k will never occur]. That the earlier $|\psi_0\rangle$ is replaced by the later $|s^k\rangle$ in the case of outcome M^k is the idea of "wave function collapse," a confusing notion best replaced with the second equality in Eq. (50).

Discussions of measurements are sometimes based on a generalization of (47) in which for any $|\psi\rangle$ in $\mathcal{H}_s$ the isometry is assumed to be of the form

$$J|\psi\rangle = \sum_j K^j|\psi\rangle \otimes |\Phi^j\rangle, \tag{51}$$

where the $\{\Phi^j\}$ are an orthonormal collection, and the *Kraus operators* K^j (note that j is a label) are arbitrary maps of $\mathcal{H}_s$ to itself subject only to the condition that

$$\sum_j (K^j)^\dagger K^j = I_s, \tag{52}$$

which guarantees that J in Eq. (51) is an isometry. Regarded as a measurement, which is to say something that determines the property of the particle at t_1, this is equivalent to a POVM in which

$$Q^j = (K^j)^\dagger K^j. \tag{53}$$

The nondestructive model in Eq. (47) is easily extended to a general PDI $\{P^j\}$ on $\mathcal{H}_s$ by setting the Kraus operator K^j in Eq. (51) equal to P^j, whence it follows that any initial $|\psi\rangle$ in $\mathcal{H}_s$ with the property P^k, i.e., $P^k|\psi\rangle = |\psi\rangle$ will result in a measurement outcome M^k and $|\psi\rangle$ will emerge unchanged at time t_2. This is the essence of Lüders' proposal [22,23], which is best regarded as a particular model of a nondestructive measurement and not (as sometimes supposed) a general principle of quantum theory.

In the case of a preparation one is not interested in the property of the particle at an earlier time, but instead its state at a time t_2 after the interaction with the measuring device is over. If, for example, the isometry is given by (47), then according to (50) if the pointer is in position k at time t_2 one can be certain that the particle is in state $[s^k]$ at this time. But a simpler and more general preparation model is obtained if in place of (47) one assumes there is a normalized state $|\psi_1\rangle$ at time t_1 and an isometry J such that

$$J|\psi_1\rangle = \sum_k \sqrt{p_k}\,|\hat{s}^k\rangle \otimes |\Phi^k\rangle, \quad M^k|\Phi^{k'}\rangle = \delta_{kk'}|\Phi^k\rangle, \tag{54}$$

where the p_k are probabilities that sum to 1. The states $|\hat{s}^k\rangle$ are normalized, but we do not assume that they form a basis; in particular, they need not be mutually orthogonal. Nonetheless one can infer that if at t_2 the pointer is in position k, the particle at this time is in the state $|\hat{s}^k\rangle$. Note that even if the $|\hat{s}^k\rangle$ are not orthogonal the states $|\hat{s}^k\rangle \otimes |\Phi^k\rangle$ are orthogonal and hence distinct; see Chap. 14 in Ref. [8] for some discussion of states of this sort. One might worry that this preparation model is stochastic: if outcome $k = 3$ is desired, sometimes it will occur and sometimes it will not. But since the pointer position is macroscopic it is not difficult to design a system whereby undesired outcomes are removed (e.g., run the particle into a barrier), or if one is repeating the experiment many times, simply keep a record of the value of k for each run, and throw out the runs for which it is not equal to 3.

D. Some remarks about density operators

The foregoing discussion of measurement models employed pure states and projectors on pure states, and it is natural to ask what the appropriate formulation ought to be if one is dealing with mixed states. Mixed states arise in quantum mechanics in two somewhat different ways. The first is analogous to a classical probability distribution: one has in mind some collection of pure states $|\psi^j\rangle$ with associated probabilities p_j, known as an *ensemble*, and the associated density operator is

$$\rho = \sum_j p_j[\psi^j]. \tag{55}$$

Suppose particles are prepared in states chosen from this ensemble with the specified probabilities, and then measured. What can one infer about the state of a particle just before the measurement, given a particular outcome? Since the only role of the initial state $|\psi_0\rangle$ in Secs. IV A and IV B is to assign probabilities, in the case of a random input one replaces $|\psi_0\rangle$ by ρ when computing averages; e.g., $\langle\psi_0|Q^k|\psi_0\rangle$ in Eq. (45) is replaced with $\mathrm{Tr}(\rho\, Q^k)$. Note that the state inferred in this way from the measurement outcome in a particular run need not be the same as the member of the ensemble sent into the measurement apparatus. This is no more surprising than the fact that the $[s^k]$ inferred in Eq. (38) can be different from $[\psi_0]$.

The second way in which a density operator arises is through taking a partial trace of an entangled pure state on a composite system down to one of the subsystems; see Chap. 15 of [8] for further details. If one is only concerned with properties of this particular subsystem and not its correlations with the others, and if only this subsystem interacts with the measuring apparatus, then the previous discussion applies: the situation is exactly the same as for the case of an ensemble. If, however, one is interested in correlations with the another subsystem or subsystems it is best to treat the entire system under consideration as a single system when working out what one can infer from a measurement, even if the measurement is carried out on just one of the subsystems, as the density operator may not provide the sort of information one is interested in. See Sec. V F below for an example.

One may also be concerned about using a pure initial state $|\Omega_0\rangle$ for a macroscopic apparatus rather than a density operator or a projector onto a large (macroscopic) subspace. This gives rise to a different set of concerns, and we refer the reader to the treatment in Chap. 17 of [8].

V. APPLICATIONS

Various applications below will illustrate the approach outlined in Sec. IV. Those in Secs. V A and V B show how a proper application of quantum principles can give physically reasonable results for the cases considered in Secs. II A and II B, while avoiding paradoxes. Simple examples of POVMs and weak measurements are considered in Secs. V C and V D. Quantum (non)contextuality and aspects of the

Einstein-Podolsky-Rosen (EPR) paradox are examined in Secs. V E and V F.

A. Spin half

The simplest nontrivial example of a quantum system is the spin of a spin-half particle, and the spin was first measured in the Stern-Gerlach experiment mentioned in every textbook. Using the notation for the eigenstates of the z component of angular momentum S_z introduced earlier in Sec. III A, suppose that a measurement of S_z corresponds to an isometry

$$J|z^j\rangle = |\Phi^j\rangle \tag{56}$$

of the form (21), where $j = +$ or $-$, and the macroscopic outcomes correspond to projectors M^+ and M^- on pointer subspaces satisfying (17). Then (24) takes the form

$$[z^+] = J^\dagger M^+ J, \quad [z^-] = J^\dagger M^- J. \tag{57}$$

Hence if the macroscopic outcome is M^+—e.g., an atom is detected in the upper beam emerging from a Stern-Gerlach magnet—one can conclude using the family of four histories at times $t_0 < t_1 < t_2$ (at t_1 and t_2 choose one of the two properties inside the curly brackets)

$$[\psi_0] \otimes [\Omega_0] \odot \{[z^+],[z^-]\} \odot \{M^+,M^-\}, \tag{58}$$

that at time t_1 before the measurement began the particle had the property $[z^+]$ corresponding to $S_z = +1/2$, whatever the initial state $[\psi_0]$. Similarly, M^- would indicate $S_z = -1/2$ at the earlier time.

One can check this by a direct calculation assuming an initial state

$$|\psi_0\rangle = \alpha|z^+\rangle + \beta|z^-\rangle, \tag{59}$$

and using the chain kets to evaluate the probabilities for the four histories in Eq. (58):

$$\begin{aligned} &\Pr(z_1^+, M_2^-) = \Pr(z_1^-, M_2^+) = 0,\\ &\Pr(z_1^+, M_2^+) = |\alpha|^2, \quad \Pr(z_1^-, M_2^-) = |\beta|^2. \end{aligned} \tag{60}$$

The marginals and conditionals are then

$$\begin{aligned} &\Pr(M_2^+) = |\alpha|^2, \quad \Pr(M_2^-) = |\beta|^2,\\ &\Pr(z_1^+ \mid M_2^+) = 1, \quad \Pr(z_1^- \mid M_2^-) = 1, \end{aligned} \tag{61}$$

where the last two hold if $|\alpha|^2$ (respectively, $|\beta|^2$) is nonzero. In short, the particle at t_1 had the value of S_z indicated by the measurement outcome at t_2, independent of the state $|\psi_0\rangle$ at t_0, in agreement with (57).

Next, assuming the same unitary dynamics (56), consider a different family of histories,

$$[x^+] \otimes [\Omega_0] \odot \{[x^+],[x^-]\} \odot \{M^+,M^-\}, \tag{62}$$

in which the initial $[\psi_0]$ is now $[x^+]$, and the properties at t_1 refer to S_x instead of S_z. It is straightforward to show that the family is consistent, with joint probabilities (obtained from chain kets)

$$\begin{aligned} &\Pr(x_1^+, M_2^+) = \Pr(x_1^+, M_2^-) = 1/2,\\ &\Pr(x_1^-, M_2^+) = \Pr(x_1^-, M_2^-) = 0. \end{aligned} \tag{63}$$

The conditionals

$$\begin{aligned} &\Pr(x_1^+ \mid M_2^+) = \Pr(x_1^+ \mid M_2^-) = 1,\\ &\Pr(x_1^- \mid M_2^+) = \Pr(x_1^- \mid M_2^-) = 0 \end{aligned} \tag{64}$$

are exactly the same for M^+ and M^-, so the measurement outcomes at t_2 tell us nothing at all about S_x at time t_1. Instead its value is determined entirely by the initial state $|x^+\rangle$ at t_0.

Given the family (62) and a pointer outcome, say M^- at t_2, are we to infer $S_x = +1/2$ at the earlier time t_1 using (64), or $S_z = -1/2$ using (61)? Both inferences are correct, *but in separate frameworks which cannot be combined*. Frameworks are chosen by the physicist depending on which aspect of the situation is of interest. The physicist who sets up an apparatus to prepare a spin-half particle with a particular polarization may wish to explain in quantum mechanical terms how it functions, in which case the family (62) is an appropriate starting point, and (64) will confirm that later measurements do not have any undesirable retrocausal influence. On the other hand the physicist who has constructed an apparatus to measure a particular polarization can best explain how it functions in that capacity by using the family (58). Even if $[\psi_0] = [x^+]$ is not an eigenstate of S_z, (61) shows that the later pointer position reveals the prior property the instrument was designed to measure. These two physicists might be one and the same; several incompatible frameworks may be useful for analyzing a particular experimental arrangement, while the single-framework rule prevents drawing meaningless conclusions or paradoxical results.

Properties at an additional intermediate time before the measurement has begun, say $t_{1.1}$, can be added to (62) to form a consistent family at times $t_0 < t_1 < t_{1.1} < t_2$, (62):

$$[x^+] \otimes [\Omega_0] \odot \{[x^+],[x^-]\} \odot \{[z^+],[z^-]\} \odot \{M^+,M^-\}, \tag{65}$$

where we assume that $T(t_{1.1},t_1) = I$. Using it one can show that

$$\Pr(x_1^+) = 1, \quad \Pr(z_{1.1}^+ \mid M_2^+) = \Pr(z_{1.1}^- \mid M_2^-) = 1. \tag{66}$$

Thus if the later measurement outcome is M^- one can be sure (based on the initial state) that $S_x = +1/2$ at t_1 and also (based on the measurement outcome) that $S_z = -1/2$ at $t_{1.1}$. This seems odd if one tries to imagine a physical process rotating the direction of the spin from $+x$ to $-z$, since the particle is moving in a field-free region and not subject to a torque. Once again the *choice of framework* which allows a description of a particular aspect of the situation must be carefully distinguished from a *dynamical physical process*. While there is no exact classical counterpart of a framework choice, the following analogy may help: If one looks at a coffee cup from above one can discern certain things—is it filled with coffee?—which are not visible from below, whereas things visible from below, such as a crack in the bottom, may not be visible from above. Changing the point of view does not change the coffee cup or its contents, but does allow one to see different things. The analogy with the quantum case breaks down in that it makes sense to speak of a cup that both contains coffee and has a (small) crack in the bottom, whereas $S_x = +1/2$ AND $S_z = -1/2$ is meaningless, as the projectors do not commute. To be sure, $S_x = +1/2$ at an earlier time is correctly combined in Eq. (65) with S_z at a later time: think of first looking at the coffee cup from the top and later from the bottom. However, interchanging the intermediate events in Eq. (65) so that S_z properties at t_1 precede the S_x properties at $t_{1.1}$ results in an inconsistent family. Classical analogies help, but in the end there is no substitute for a consistent quantum analysis.

B. Mach-Zehnder

A correspondence between spin-half measurements as discussed in Sec. V A and the Mach-Zehnder setup of Sec. II B will assist in understanding the latter. Consider a time t_1 at which, see Fig. 3, the photon has been reflected from the upper and lower mirrors, but has yet to reach the location of the second beam splitter, or, if the latter is absent, the crossing point of the two trajectories. Let $|z^+\rangle$ be the part of the photon wave packet in the upper arm, and $|z^-\rangle$ the part in the lower arm of the interferometer at this time, and let $|x^+\rangle$ and $|x^-\rangle$ be the coherent superpositions of $|z^+\rangle$ and $|z^-\rangle$ defined in Eq. (4). Further assume that the action of the first beam splitter in Fig. 3 is to prepare the photon in the state $|x^+\rangle$. Let M^+ be the projector on the macroscopic subspace in which D^+ in Fig. 3 has detected the photon while D^- has not, and M^- its counterpart for detection by D^- rather than D^+.

If the second beam splitter is absent, Fig. 3(b), a photon in the state $|z^+\rangle$ in the upper arm will trigger D^+, while $|z^-\rangle$ in the lower arm will trigger D^-. This can be discussed using a family of four histories as in Eq. (58), with $|\psi_0\rangle = |x^+\rangle$:

$$[x^+] \otimes [\Omega_0] \odot \{[z^+],[z^-]\} \odot \{M^+,M^-\}. \quad (67)$$

The conditional probabilities are the same as in Eq. (61): if D^+ is triggered one can be certain the photon was earlier in the state $[z^+]$, so in the upper arm of the interferometer, whereas detection by D^- indicates the earlier state $[z^-]$ in the lower arm. These are the same conclusions one would arrive at from a naive inspection of Fig. 3(b), but they have now been confirmed using an analysis based on consistent quantum principles.

Now add an additional time $t_{1.1} > t_1$ at which the photon is still inside the interferometer. The consistent family

$$[x^+] \otimes [\Omega_0] \odot \{[x^+],[x^-]\} \odot \{[z^+],[z^-]\} \odot \{M^+,M^-\} \quad (68)$$

(where note that histories with $[x^-]$ at t_1 have zero probability, so can be ignored) is formally identical to (65), but introduces a new conceptual difficulty. In the spin-half case the issue was how a spin angular momentum of $S_x = +1/2$ at t_1 could suddenly precess into $S_z = +1/2$ or $-1/2$ at $t_{1.1}$. However mysterious that might be, one could still imagine the change taking place at the location of the spin-half particle. But for the Mach-Zehnder $[x^+]$ is a nonlocal superposition between the two arms at t_1; can it suddenly collapse into one or the other arm, $[z^+]$ or $[z^-]$, at a time $t_{1.1}$, even if the interval between t_1 and $t_{1.1}$ is very short, so making this collapse essentially instantaneous? Is this (seeming) nonlocality consistent with relativity theory?

Just as in the case of spin half this (apparent) paradox may be dealt with by noting that a change in what is being described is not the same as a physical process. Thus if the pair $\{[z^+],[z^-]\}$ at $t_{1.1}$ in Eq. (68) is replaced with $\{[x^+],[x^-]\}$, this new family is again consistent, but the "collapse" between t_1 and $t_{1.1}$ is no longer present. Families or frameworks are chosen by the physicist and are not consequences of some law of nature. See the discussion following (66).

In addition it is worth noting that a quantum superposition, such as $|x^+\rangle$, of a particle at two locations is not at all the same thing as its being in both places at the same time. Translated into quantum terminology the statement that the photon is in the upper arm AND in the lower arm becomes $[z^+]$ AND $[z^-]$, which because the projectors commute makes perfectly good sense, and because their product is zero this conjunction is always false: a photon can *never* be located simultaneously in both the upper and in the lower arm, unlike a classical wave.

Next consider the case with the second beam splitter present, and suppose that the phases are such that a photon in the state $|x^+\rangle$ inside the interferometer will later be detected by D^+, and $|x^-\rangle$ detected by D^-. In this case one can think of the detectors and the second beam splitter as together constituting an apparatus designed to detect $[x^+]$ and $[x^-]$, the photon analogs of the spin-half S_x eigenstates (which for a spin-half particle could be measured by rotating the Stern-Gerlach apparatus so that its field gradient is in the x rather than the z direction). The second beam splitter changes the unitary dynamics in such a way that $|z^j\rangle$ on the left side of (56) is replaced by $|x^+\rangle$, and thus (57) becomes

$$[x^+] = J^\dagger M^+ J, \quad [x^-] = J^\dagger M^- J. \quad (69)$$

Thus the measurement outcomes now indicate different superposition states of the photon inside the interferometer; the measurement measures "which phase?" rather than "which path?"; see Fig. 4. With the new dynamics (67) is no longer a consistent family, but one can instead use

$$[\psi_0] \otimes [\Omega_0] \odot \{[x^+],[x^-]\} \odot \{M^+,M^-\}, \quad (70)$$

in order to infer from the measurement outcome the presence of one of two distinct (i.e., orthogonal) superposition states inside the interferometer, which is to say the difference between photons originating in S_1 or S_2 in Fig. 4(a).

We are now ready to discuss the (supposed) paradox, Sec. II B, associated with removing or inserting the second beam splitter at the very last moment just before the photon reaches it. One can think of this change as the Mach-Zehnder analog of rotating a Stern-Gerlach apparatus about the axis of the atomic beam just before the arrival of a spin-half particle, so that it will measure S_x rather than S_z. In the absence of this rotation one can use the measurement outcome to assign a value to S_z before the measurement took place, whereas if the rotation has taken place before the particle arrives, the measurement outcome indicates the prior value of S_x. This does not mean that the particle has both an S_z and an S_x value, for these two quantities are incompatible, and that is why they cannot be measured simultaneously. In the case of the Mach-Zehnder, if the second beam splitter is absent the measurement outcome will indicate either an earlier $[z^+]$ property, photon in the upper arm, or $[z^-]$, photon in the lower arm. If the second beam splitter is present the measurement outcome distinguishes the earlier superposition properties $[x^+]$ and $[x^-]$, neither of which is compatible with assigning the photon to one of the arms rather than the other. In neither case is there any need to suppose that the later measurement choice influences the particle before measurement. Instead, changing the type of measurement alters what type of information about the earlier state of the particle can be inferred from the measurement outcome.

C. Spin-half POVM

A simple but instructive example of a POVM for a spin-half particle can be constructed using the three nonorthogonal states

$$|u^1\rangle = (|z^+\rangle + |z^-\rangle)/\sqrt{2}, \quad |u^2\rangle = (\omega|z^+\rangle + \omega^2|z^-\rangle)/\sqrt{2},$$
$$|u^3\rangle = (\omega^2|z^+\rangle + \omega|z^-\rangle)/\sqrt{2}, \quad \omega := \exp[2\pi i/3]. \tag{71}$$

The projectors $[u^k]$ are associated with points on the equator of the Bloch sphere: $[u^1]$ at the positive x axis, while $[u^2]$ and $[u^3]$ are separated from $[u^1]$ and each other by 120°. The operators

$$Q^k := (2/3)[u^k] \tag{72}$$

for $k = 1,2,3$ sum to the identity and hence constitute a POVM.

This POVM can be obtained from an isometry as discussed in Sec. IV B, where we assume for simplicity a "toy" apparatus Hilbert space $\mathcal{H}_M$ of dimension 3, with an orthonormal basis $\{|k\rangle\}$, $k = 1,2,3$. The isometry J can be written in the form

$$J|u^k\rangle = |v^k\rangle := \sqrt{3/2}\,|k\rangle - \sqrt{1/2}\,|w\rangle,$$
$$|w\rangle := (|1\rangle + |2\rangle + |3\rangle)/\sqrt{3}. \tag{73}$$

This J maps the two-dimensional $\mathcal{H}_s$ into the two-dimensional subspace of $\mathcal{H}_M$ consisting of kets orthogonal to $|w\rangle$. The orthogonal measurement projectors in the notation of Sec. IV are

$$M^k := [k]. \tag{74}$$

With the help of the formulas

$$|u^k\rangle = J^\dagger|v^k\rangle, \quad J^\dagger|w\rangle = 0, \quad |k\rangle = \sqrt{2/3}\,|v^k\rangle + \sqrt{1/3}\,|w\rangle, \tag{75}$$

where the first and second are consequences of $J^\dagger J = I_s$ and the fact that $\langle w|J|u^k\rangle = 0$, while the third comes from rewriting (73), one can show that

$$J^\dagger M^k J = J^\dagger[k]J = (2/3)[u^k] = Q^k, \tag{76}$$

in agreement with (35).

The analysis in Sec. IV B shows that the family consisting of the histories

$$Y^k = [\psi_0] \otimes [\Omega_0] \odot \{[u^k], I_s - [u^k]\} \odot M^k \tag{77}$$

at times $t_0 < t_1 < t_2$, $k = 1,2,3$, is consistent for any initial spin-half state $|\psi_0\rangle$. Note that the PDI $\{[u^k], I_s - [u^k]\}$ at t_1 is linked to the final M^k, and because the $[u^k]$ are not orthogonal these intermediate PDIs for different k are incompatible. This is not a problem, because in a particular run only one measurement outcome corresponding to a specific k will occur, and for *that* k one can be sure (see the discussion in Sec. IV B; here Q^k is proportional to a rank-one projector) the particle was at time t_1 in the state $[u^k]$, since the history with the event $I_s - [u^k]$ has zero weight.

That the framework used to describe the situation at t_1 depends on the later measurement outcome at time t_2 should not be misunderstood. It is not the case that a later event *caused* an earlier one. Rather, a specific later outcome of a process which is intrinsically random allows one to reach a conclusion which would not have been possible had the outcome been different. There are classical analogs of this, though of course they all have limitations when discussing quantum systems. One should also keep in mind that while the family (77) provides a rather natural interpretation of the measurement outcome k, the choice is not unique.

It is worth considering what happens if one is trying to calibrate the POVM apparatus using several runs in which particles are prepared in the state $[u^1]$, i.e., $S_x = 1/2$. The probability of outcome k will be

$$\Pr\left(M_2^k\right) = \mathrm{Tr}([u^1]\,Q^k) = \begin{cases} 2/3, & \text{if } k = 1,\\ 1/3, & \text{if } k = 2 \text{ or } 3. \end{cases} \tag{78}$$

Thus unlike the situation for a PDI, the prior preparation does not determine the measurement outcome, although M^1 is more likely to occur than either of its alternatives. If for some run the outcome is M^2 we might conclude, using (77) with $k = 2$, that the particle was earlier in the state $[u^2]$, even though we know it was prepared in $[u^1]$. This is not a paradox as long as one remembers that quantum theory allows the use of different frameworks, and one is careful not to combine incompatible frameworks in violation of the single-framework rule. An alternative calibration procedure uses particles prepared in states orthogonal to the $[u^k]$. For example, $[x^-]$ is orthogonal to $[u^1]$, and if in Eq. (77) $|\psi_0\rangle = |x^-\rangle$, the outcome probabilities are

$$\Pr\left(M_2^k\right) = \mathrm{Tr}(\,[x^-]\,Q^k) = \begin{cases} 0, & \text{if } k = 1,\\ 1/2, & \text{if } k = 2 \text{ or } 3. \end{cases} \tag{79}$$

The fact that in this case the $k = 1$ outcome is never observed is an indication that the apparatus is functioning properly.

D. Weak measurements

A *weak measurement* is one in which the measured system, the particle, interacts very weakly with the measurement apparatus. As a consequence a single measurement provides very little information about the particle, so weak measurements are usually employed in a situation in which the measurement can be repeated many times, each time with a particle prepared in the same state before the measurement. One way of implementing a weak measurement is to let the particle interact weakly with a another microscopic system, called a probe, which has itself been prepared in a known quantum state. After interacting with the particle the probe is subjected to a projective ("strong") measurement by a macroscopic apparatus, with the intent of learning something about the original particle in this indirect way. There are many possible variations of this procedure. For example, the same particle may be subjected to a succession of weak measurements, one after the other, each supplying some additional information. Or, after interacting with the probe, the particle may itself be subjected to a strong measurement. When attention is focused on cases resulting in some particular outcome of the final strong measurement on the particle one speaks of *postselection*. There is an enormous literature on weak measurements; for access to some of it see [24–26].

Weak measurements have no necessary connection with *weak values*, though the two are often discussed together, and sometimes it is assumed that weak measurements can or should be be interpreted using weak values. The physical significance of weak values has been the subject of an ongoing controversy [25]. Suffice it to say that in general there is no reason to think of a weak value as linked to a physical property, or the average value of a physical variable, at least as those terms are employed in the present article, where they are associated with Hilbert subspaces.

A weak measurement, either by itself or when followed by a strong measurement, can be understood as a particular type of POVM, and thus understood in terms of prior properties of the particle as discussed in Sec. IV B. The following simple example illustrates how this works in a particular case. Let the particle be a two-state system with an orthonormal basis $\{|A\rangle,|B\rangle\}$, which one can think of as representing a photon in one of two channels, as in the Mach-Zehnder interferometer considered earlier in Secs. II B and V B. The three-dimensional Hilbert space $\mathcal{H}_r$ for the probe has an orthonormal basis $\{|j\rangle\}, j = 0,1,2$. We assume the probe is initially in the state $|0\rangle$, while the particle is in a superposition

$$|\psi_0\rangle = a|A\rangle + b|B\rangle. \tag{80}$$

The particle-probe interaction results in a unitary time development

$$\begin{aligned} T(t_2,t_1)(|A\rangle \otimes |0\rangle) &= |A\rangle \otimes (\zeta|0\rangle + \eta|1\rangle),\\ T(t_2,t_1)(|B\rangle \otimes |0\rangle) &= |B\rangle \otimes (\zeta|0\rangle + \eta|2\rangle) \end{aligned} \tag{81}$$

during the interval from t_1 to t_2 [or t_0 to t_2, given our usual assumption that $T(t_2,t_0) = T(t_2,t_1)$], where

$$\zeta = \sqrt{1-\epsilon}, \quad \eta = \sqrt{\epsilon}. \tag{82}$$

Here ϵ, a measure of the strength of the particle-probe interaction, is assumed to be very small, so that the probability is high that the probe will be left in its initial untriggered state $|0\rangle$, but on rare occasions it will be kicked to $|1\rangle$ if the particle is in channel A, or to $|2\rangle$ if the particle is in B. Feynman's use of a weak light source in Sec. 1-6 of [12] is a good illustration of this idea.

After this, during the time interval from t_2 to t_3 the probe is measured in the $j = 0,1,2$ basis, and the particle is measured in an orthonormal basis $|E\rangle,|F\rangle$ related to $|A\rangle,|B\rangle$ by

$$|A\rangle = \alpha_e|E\rangle + \alpha_f|F\rangle, \quad |B\rangle = \beta_e|E\rangle + \beta_f|F\rangle, \tag{83}$$

where

$$\begin{pmatrix} \alpha_e & \alpha_f \\ \beta_e & \beta_f \end{pmatrix} \tag{84}$$

is a unitary matrix. Different choices of these parameters could be used to represent different situations analogous to those shown in Fig. 3, where the second beam splitter is either present or absent.

As the particle and the probe are measured by separate devices we can associate with each an isometry mapping from t_2 to t_3:

$$\begin{aligned} J_s|E\rangle &= \left|\Phi_s^E\right\rangle, \quad J_s|F\rangle = \left|\Phi_s^F\right\rangle,\\ J_r|j\rangle &= \left|\Phi_r^j\right\rangle \quad \text{for } j = 1,2,3. \end{aligned} \tag{85}$$

Combining these with (81) yields an isometry mapping $\mathcal{H}_s$ at time t_1 to both outputs at time t_3:

$$\begin{aligned} J|A\rangle &= \zeta(\alpha_e|\Phi^{E0}\rangle + \alpha_f|\Phi^{F0}\rangle) + \eta(\alpha_e|\Phi^{E1}\rangle + \alpha_f|\Phi^{F1}\rangle),\\ J|B\rangle &= \zeta(\beta_e|\Phi^{E0}\rangle + \beta_f|\Phi^{F0}\rangle) + \eta(\beta_e|\Phi^{E2}\rangle + \beta_f|\Phi^{F2}\rangle), \end{aligned} \tag{86}$$

where $|\Phi^{E0}\rangle$ is shorthand for $|\Phi_s^E\rangle \otimes |\Phi_r^0\rangle$, and lies in the subspace $M^{E0} = M^E \otimes M^0$ for the two pointers, and similarly for the other cases.

The backward $J^\dagger(\cdot)J$ map applied to the pointer projectors yields POVM elements which are operators on $\mathcal{H}_s$ and thus can be written as 2×2 matrices in the $|A\rangle,|B\rangle$ basis:

$$\begin{aligned} Q^{E0} &= (1-\epsilon)\begin{pmatrix} |\alpha_e|^2 & \alpha_e^*\beta_e \\ \alpha_e\beta_e^* & |\beta_e|^2 \end{pmatrix},\\ Q^{F0} &= (1-\epsilon)\begin{pmatrix} |\alpha_f|^2 & \alpha_f^*\beta_f \\ \alpha_f\beta_f^* & |\beta_f|^2 \end{pmatrix},\\ Q^{E1} &= \epsilon\begin{pmatrix} |\alpha_e|^2 & 0 \\ 0 & 0 \end{pmatrix}, \quad Q^{E2} = \epsilon\begin{pmatrix} 0 & 0 \\ 0 & |\beta_e|^2 \end{pmatrix},\\ Q^{F1} &= \epsilon\begin{pmatrix} |\alpha_f|^2 & 0 \\ 0 & 0 \end{pmatrix}, \quad Q^{F2} = \epsilon\begin{pmatrix} 0 & 0 \\ 0 & |\beta_f|^2 \end{pmatrix}. \end{aligned} \tag{87}$$

That these six operators sum to the identity follows from the unitarity of (84): its rows are orthogonal, so $\alpha_e^*\beta_e + \alpha_f^*\beta_f = 0$, and its column vectors are normalized.

The simple form of the last four matrices in Eq. (87) can be understood by noting that the probe, which starts off in $|0\rangle$, can reach state $|1\rangle$ only if the particle is in channel A, which is why Q^{E1} and Q^{F1} are proportional to the projector $[A]$; similarly, only if the particle is in B can the probe arrive at $|2\rangle$. The more complicated matrix Q^{E0} is $1-\epsilon$ times the projector $[E]$, which is reasonable since in this case the probe was not triggered but remained in $|0\rangle$, so did not perturb the particle; similarly, $Q^{F0} = (1-\epsilon)[F]$.

Our discussion has employed the strategy introduced in Sec. IV B, of interpreting outcome k of a generalized measurement in terms of properties that correspond to diagonalizing the operator Q^k. In this example each Q^k is proportional to a pure state projector, so the interpretation is relatively simple, and is independent of the initial state $|\psi_0\rangle$ of the particle in Eq. (80). The *probabilities* of various measurement outcomes will depend on the coefficients a and b in $|\psi_0\rangle$, and can be computed from the POVM matrices using $\langle\psi_0|Q|\psi_0\rangle$, whereas the *physical interpretation* of each outcome in terms of prior properties does not depend on $|\psi_0\rangle$. The framework used here is convenient for discussing what a quantum measurement measures, but does not exclude the use of other frameworks. The standard textbook computational procedure uses the entangled state $T(t_2,t_1)(|\psi_0\rangle \otimes |0\rangle)$ to calculate probabilities of various measurement outcomes, and there is nothing wrong with that when one is only interested in those probabilities, and not in how these outcomes reveal the properties of the particle that the apparatus was designed to measure.

For an analogous discussion (without using the language of POVMs) of a more complicated situation, which has given rise to some controversy, see Sec. V of [2].

E. Is quantum mechanics contextual?

One often encounters the claim that "quantum mechanics is contextual."[2] Unfortunately the term "contextual" is used in more than one way. A relatively precise definition due to Bell [28] and used in some later quantum foundations literature, e.g., Sec. VII of [29] and p. 188 of [30], is the following: Let A, B, and C be three observables (i.e., Hermitian operators), and suppose that A commutes with B and C, but B and C do not commute:

$$[A,B]=0, \quad [A,C]=0, \quad [B,C]\neq 0. \tag{88}$$

This means that A can (in principle) be measured together with B, or together with C, whereas B and C are incompatible and cannot be measured together. One can then ask, does the measured value of A depend on whether it is measured together with B or with C? If the answer is "yes," then quantum mechanics (or whatever theory is being discussed) is *contextual*, and if "no," it is *noncontextual*. To avoid confusion, let us add a modifier and refer to *Bell (non)contextual* when these terms are used in the way just described. The following argument will show that quantum mechanics in the consistent histories interpretation is *Bell noncontextual*. (A more recent and somewhat different definition of "contextual" is discussed briefly at the end of this section.)

The definition given above runs into the following difficulty. In a single experimental run A cannot be measured together with both B and C, since B and C cannot be measured in the same run. And the measured value of A may vary randomly from run to run, making it difficult to make a comparison between those in which A is measured with B and those in which it is measured together with C. Let us explore this difficulty by thinking of an apparatus equipped with a switch with two settings: β and γ. With the switch at β the apparatus will measure both A and B, while with the setting γ it will measure A and C. We suppose that the apparatus has been calibrated, Sec. II C, for A measurements for both switch settings, so the experimenter can be reasonably confident that the A pointer outcome will give the correct answer if the input state is an eigenstate of A. Similarly, B calibrations can be carried out with the switch at β, and C calibrations for the γ setting.

Now we ask, Suppose that with the β setting the A measurement outcome corresponds to a particular eigenvalue, say a_3. Would this outcome have been the same if the switch setting had been γ? Counterfactual questions of this sort are a bit tricky; see [31] and Chap. 19 of [8] for a proposal that gives plausible results in a quantum setting. For the present discussion the basic idea is that if one can reliably infer from the apparatus outcome with switch setting β the eigenvalue of A that characterized the particle *before* any interaction with the apparatus, it seems reasonable that changing the switch from β to γ at the very last moment could not have altered that earlier property, so the result would have been the same with the γ setting, given that the apparatus had been calibrated.

[2]Some authors make it clear that it is *hidden variables* versions of quantum mechanics which are contextual, but many omit that qualification; for a recent (but hardly unique) example, see [27].

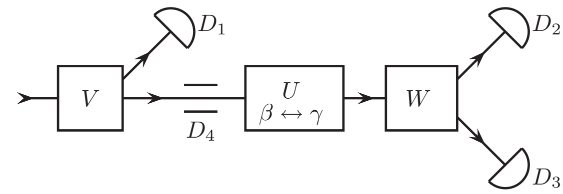

FIG. 5. Apparatus to measure A along with B (U_β), or with C (U_γ).

To make things less abstract, consider a spin-one particle, and let $|1\rangle$, $|2\rangle$, $|3\rangle$ be an orthonormal basis for its Hilbert space $\mathcal{H}_s$. Define the following observables using dyads:

$$\begin{aligned} A &= |1\rangle\langle 1| - |2\rangle\langle 2| - |3\rangle\langle 3|,\\ B &= \tfrac{1}{2}|1\rangle\langle 1| + |2\rangle\langle 2| - |3\rangle\langle 3|,\\ C &= 2|1\rangle\langle 1| + |2\rangle\langle 3| + |3\rangle\langle 2|. \end{aligned} \tag{89}$$

It is obvious that $[A,B]=0$, and straightforward to show that $[A,C]=0$ and $[B,C]\neq 0$.

A possible apparatus for measuring these observables is shown schematically in Fig. 5. The incoming particle first passes through a device V (one can think of an electric field gradient acting on a particle with an electric quadrupole moment) which splits the path in two. The upper path is followed by a particle in the state $|1\rangle$ and leads to the detector D_1. The lower (straight) path is followed by a particle whose state is any linear combination of $|2\rangle$ and $|3\rangle$, and it passes through a nondestructive detector D_4 that measures the particle's passage without disturbing its internal state. Following this there is another device U with a switch: if the switch setting is β it carries out a unitary transformation U_β equal to the identity I (i.e, the device does nothing), while if the setting is γ the unitary is

$$U_\gamma = (1/\sqrt{2})\{|2\rangle\langle 2| + |2\rangle\langle 3| + |3\rangle\langle 2| - |3\rangle\langle 3|\}. \tag{90}$$

Then yet another device W (think of a Stern-Gerlach magnet) splits the trajectory into one moving upwards if the particle state is $|2\rangle$, or downwards if it is $|3\rangle$; these terminate in detectors D_2 and D_3.

A particle initially in the eigenstate $|1\rangle$ of A with eigenvalue $+1$ will be detected by D_1, whereas any eigenstate of A with eigenvalue -1, i.e., any linear combination of $|2\rangle$ and $|3\rangle$, will be detected by D_4 and then travel on. Thus a measurement of A precedes the particle's passing through the box U, and the outcome will not be affected by whether the unitary is U_β or U_γ. The switch setting could, in principle, be decided at the very last moment, after the particle (if on this path) has passed through D_4. A measurement of B is carried out by setting $U=I$, so that initial eigenstates with eigenvalues of $1/2$, 1, and -1 will be detected by detectors 1, 2, and 3, respectively. Alternatively, C can be measured by setting $U=U_\gamma$, (90), and its eigenvalues of 2, 1, and -1 correspond to detection by detectors 1, 2, and 3, respectively. It should be clear from the construction shown in Fig. 5 that if the change from an A-plus-B apparatus to an A-plus-C apparatus, by moving the switch from β to γ, is made *after* the particle has passed the position of detectors D_1 and D_4, this cannot affect the A measurement outcome, assuming the future does not influence the past. Thus in this case it seems evident that the measurement is (Bell) noncontextual.

The preceding discussion for a particle with three states is easily generalized to the case of an arbitrary (finite) number of states. To see this, let $\{P^j\}$ be the PDI that diagonalizes A with different projectors associated with different eigenvalues; i.e.,

$$A = \sum_{\alpha} a_j P^j, \tag{91}$$

and $a_j \neq a_{j'}$ when $j \neq j'$. Then it is straightforward to show that if A commutes with B, every P^j in Eq. (91) also commutes with B. So if a basis is chosen such that the matrix of A is diagonal with separate blocks for each eigenvalue, the matrix of B will be block diagonal, and each of its blocks can be separately diagonalized by a change of basis that leaves the A matrix unchanged. The same comment applies to any other observable C that commutes with A, whether or not it commutes with B, though of course the bases used to diagonalize B and to diagonalize C must be different if $[B,C] \neq 0$. The V box in Fig. 5 separates incoming particles into separate beams corresponding to the different eigenvalues of A, and in each beam there is a nondestructive detector that plays the role of D_4 in Fig. 5. These measurements determine the value of A. Next in each beam there is a unitary that depends on the choice of β or γ, followed by a final set of detectors from which the eigenvalues of B or C, as the case may be, can be inferred.

This example leaves open the possibility that if the time ordering were different, B or C, as the case might be, measured *before* A, this might have an effect on the value of A. Also we have been assuming that the particle enters the apparatus in an eigenstate of A; what if it is in some arbitrary superposition state $|\psi_0\rangle$? Both concerns are easily handled using the measurement model introduced in Sec. IV. In particular, (35) takes the form

$$P^k = J^\dagger M^k J, \tag{92}$$

for a projective measurement associated with the PDI $\{P^k\}$, the obvious generalization of (24). Thus one can be certain that the particle possessed the property P^k corresponding to the eigenvalue a_k of A at the time t_1 before the measurement took place, given the later measurement outcome (pointer position) k that corresponds to M^k. What went on at an intermediate time cannot alter this, always assuming the apparatus has been properly calibrated, so that (92) holds. Hence quantum measurements carried out with a properly designed and tested apparatus are noncontextual, and in this sense quantum theory is (Bell) *noncontextual*.

So why is it that one is sometimes told, often with great confidence, that quantum theory is *contextual*? Various reasons suggest themselves. The first is that measurements are not properly treated in textbooks. One admires textbook authors (e.g., [7,32,33]) who are brave enough to agree publicly with Bell [34]: they have not been able to solve the measurement problem. And without some, at least implicit, theory of quantum measurements one cannot even begin to discuss contextuality in Bell's sense of the word. Another reason is that in attempting to fill this serious gap in the textbooks, John Bell and others have proposed that microscopic properties rather than being represented by Hilbert subspaces might correspond to hidden variables which in certain crucial respects are *classical*. This is obvious in the best-known hidden variables approach, the de Broglie–Bohm pilot wave [11,35], where a quantum particle is assumed to have a well-defined classical position at all times. But it is also true of the mysterious quantity λ that appears in many discussions of Bell inequalities. There is always an assumption of classical behavior on the part of this mythical object, as has been pointed out repeatedly by Fine, e.g., Sec. 3 of [36], and clearly comes to light in a proper quantum mechanical analysis of the situation [37]. Even when authors declare that λ is or could be the "quantum state," they are not referring to the noncommuting projectors representing quantum properties in von Neumann's sense. Decades of research on hidden variables theories have not come close to solving the second measurement problem [38–40].

Sometimes the paradoxes and associated inequalities of Kochen and Specker [41], the Mermin square (Sec. V of [29]), and the like are invoked as grounds for believing that quantum mechanics is contextual, so it is worth pointing out where such claims go astray, at least in the case of what we are calling Bell contextuality. (For a more detailed discussion, see Chap. 22 of [8].) Suppose A commutes with B. Then, see the discussion following (91), it is possible to write down a collection of pairs of eigenvalues (a_j,b_k), each pair corresponding to some well-defined and nontrivial (i.e., not just the zero vector) Hilbert subspace where A takes the value a_j and B the value b_k. Similarly, if A commutes with C one can construct a similar list (a_j,c_l) of possible joint values. One might suppose that by comparing these two lists one could find pairs (b_k,c_l) of possible joint values of B and C. In particular, suppose that (a_2,b_2) is a member of the first list. Then there would surely be at least one pair in the second list, say (a_2,c_3), such that (b_2,c_3) is a pair of possible simultaneous values for B and C. Perfectly good classical reasoning, but it can fail in the quantum case if B and C do not commute; the reader can construct an example using (89). By applying this reasoning, which violates the single-framework rule, a sufficient number of times using a sufficient number of observables one can arrive at a contradiction, and this, so it is claimed, implies that quantum mechanics is contextual. But this is not a demonstration of the Bell contextuality of quantum mechanics; instead it shows that the single-framework rule must be taken seriously if one wishes to reason in a consistent way about microscopic quantum systems.

It is worth remarking that if Bell contextuality were true this would seriously undermine quantum physics as an experimental science, since experimenters often interpret their data in terms of prior microscopic properties once the apparatus has been calibrated. And calibration refers to the quantity of interest, A in the above discussion, not to other observables which the apparatus might quite incidentally be measuring at the same time. It would be an insuperable task to take all of these other possibilities into account when designing or calibrating equipment. Thus experimental physics relies on the fact that quantum mechanics is Bell *noncontextual*.

Finally, there is an alternative definition of "contextual" that appears to underlie many of the more recent discussions in the literature, and receives a precise definition in Ref. [42]. A *context* is defined to be a collection of commuting observables which can be measured simultaneously, thus associated with a single PDI, or in consistent histories terminology a framework. In the example in Eq. (89), A and B belong to one context, and A and C to another, but there is no context (framework) which contains all three. Given some collection of contexts and a single initial quantum state, one can use the Born rule to compute the probabilities of measurement outcomes for operators in each context. The probability assigned to a particular operator A that belongs to several different contexts is independent of the context (as expected, since quantum theory is Bell noncontextual). However there may not exist a *joint* probability distribution for the *entire* collection of observables if not all of them commute, and hence there is no single context that contains them all. The *absence* of such a joint distribution is taken to indicate that quantum mechanics (or whatever theory is under consideration) is *contextual*. Perhaps "multicontextual" would be a better term.

F. Einstein-Podolsky-Rosen-Bohm

The Einstein-Podolsky-Rosen (EPR) paradox [43] is well known and has given rise to an enormous number of publications. The purpose of the following remarks is to relate it to the second measurement problem, using Bohm's simple version of EPR in Ref. [44]. It makes use of the singlet spin state

$$\sqrt{2}\,|\psi_0\rangle = |z^+\rangle_a \otimes |z^-\rangle_b - |z^-\rangle_a \otimes |z^+\rangle_b = |x^+\rangle_a \otimes |x^-\rangle_b - |x^-\rangle_a \otimes |x^+\rangle_b \tag{93}$$

in the Hilbert space $\mathcal{H}_a \otimes \mathcal{H}_b$ of two spin-half particles a and b, thought of as quite far apart so they do not interact with each other, and particle b will not interact with an apparatus carrying out a measurement on particle a.

The essence of the original EPR argument expressed using Bohm's model is as follows. A measurement of S_z for particle a can be used to infer the value of S_z for b, and since particle a and the apparatus are not interacting with b, that particle must have had that value of S_z before the measurement of a took place. The property of particle b was, so to speak, "really there," a part of physical reality. But one could just as well measure S_x for particle a, and via the same sort of argument assign a value to S_x for particle b, which again would be "really there." But in the two-dimensional Hilbert space of a spin-half particle there is nothing to represent a situation in which both S_x and S_z simultaneously take on particular values. Thus the Hilbert space approach does not provide a complete description of physical reality; something is missing.

We shall assume that only particle a is measured, and that since neither it nor the measurement apparatus can interact with particle b, the corresponding isometry J, see Sec. IV A, that relates the spin states of both particles, $\mathcal{H}_s = \mathcal{H}_a \otimes \mathcal{H}_b$, to the measurement outcome can be written in the form

$$J(|\psi\rangle_a \otimes |\chi\rangle_b) = (J_a|\psi\rangle_a) \otimes |\chi\rangle_b, \tag{94}$$

where $|\psi\rangle$ and $|\chi\rangle$ are any two elements of $\mathcal{H}_a$ and $\mathcal{H}_b$, and $J_a : \mathcal{H}_a \to \mathcal{H}_M$ is the isometry for a measurement of particle a alone. For an S_z measurement, J_a tales the form

$$J_a|z^+\rangle_a = |A^+\rangle, \quad J_a|z^-\rangle_a = |A^-\rangle, \quad M^+|A^+\rangle = |A^+\rangle, \quad M^-|A^-\rangle = |A^-\rangle, \tag{95}$$

where, as in Sec. IV A, M^+ and M^- are projectors on the macroscopic pointer position subspaces representing the possible outcomes of the measurement. The counterpart of (24) is

$$[z^k]_a \otimes I_b = J^\dagger(M^k \otimes I_b)J, \quad k = +\text{ or}-, \tag{96}$$

where I_b is the identity for particle b.

Consider a family of histories at times $t_0 < t_1 < t_2$:

$$[\psi_0] \otimes [\Omega_0] \odot \{\, [z^+]_a\,,[z^-]_a\} \otimes \{\, [z^+]_b\,,[z^-]_b\} \odot \{M^+,M^-\}, \tag{97}$$

where the four projectors $[z^+]_a \otimes [z^+]_b$, etc., at the intermediate time sum to the identity on $\mathcal{H}_a \otimes \mathcal{H}_b$. There are eight histories in this family, but we only need to pay attention to those in which $[z^+]_a$ at time t_1 is followed by M^+ at t_2, or $[z^-]_a$ by M^-, since the other chain kets vanish. But in addition, for $|\psi_0\rangle$ as defined in Eq. (93),

$$([z^+]_a \otimes [z^+]_b)|\psi_0\rangle = ([z^-]_a \otimes [z^-]_b)|\psi_0\rangle = 0. \tag{98}$$

This means that only two histories have positive probabilities: $[z^+]_a \otimes [z^-]_b$ followed by M^+ or $[z^-]_a \otimes [z^+]_b$ followed by M^-. The chain kets are obviously orthogonal, so the family is consistent, and each of these histories is assigned a probability of 1/2, leading to the conditional probabilities:

$$\Pr(z^+_{a1}\,|\,M^+_2) = \Pr(z^-_{b1}\,|\,M^+_2) = 1, \quad \Pr(z^-_{a1}\,|\,M^-_2) = \Pr(z^+_{b1}\,|\,M^-_2) = 1. \tag{99}$$

In words, the outcome M^+ of the measurement of S_z for particle a indicates that at the earlier time S_z was $+1/2$ for particle a and $-1/2$ for particle b, while M^- means S_z was $-1/2$ for a and $+1/2$ for b.

A second consistent family, using the same isometry (95) appropriate for measuring S_z, employs eigenstates of S_x rather than S_z at t_1:

$$[\psi_0] \otimes [\Omega_0] \odot \{[x^+]_a,[x^-]_a\} \otimes \{[x^+]_b,[x^-]_b\} \odot \{M^+,M^-\}. \tag{100}$$

In this case the chain kets in which $[x^+]_a$ and $[x^-]_a$ are followed by M^+ or M^- do not have to vanish. However, the initial $[\psi_0]$ eliminates histories that contain $[x^+]_a \otimes [x^+]_b$ or $[x^-]_a \otimes [x^-]_b$ at t_1, leaving only four nonzero chain kets, which are orthogonal (something the reader may wish to check). The resulting probabilities then lead to

$$\Pr(x^+_{a1} \otimes x^-_{b1}\,|\,M^+_2) = \Pr(x^+_{a1} \otimes x^-_{b1}\,|\,M^-_2) = 1/2, \quad \Pr(x^-_{a1} \otimes x^+_{b1}\,|\,M^+_2) = \Pr(x^-_{a1} \otimes x^+_{b1}\,|\,M^-_2) = 1/2. \tag{101}$$

Since these conditional probabilities are the same for the two measurement outcomes M^+ and M^-, the later measurement provides no additional information; that S_x has opposite values for particles a and b is a consequence of the initial state (93).

We are now in a position to discuss the Bohm version of EPR using a consistent theory of quantum measurements. The analysis based on the history family (97) indicates that one can, indeed, infer from a measurement of S_z on particle a the value of S_z for particle b. However, see (101), the S_z measurement of particle a tells one nothing about S_x for either particle a or particle b. To which the response might be, Make an S_x measurement on particle a, and the outcome will then tell one S_x for particle b. This is entirely correct, but of course one cannot measure both S_x and S_z for particle a, because there is nothing there to be measured, at least if one is using Hilbert space quantum mechanics. As for the counterfactual, "S_z was measured for particle a and the value was $+1/2$, but if instead S_x had been measured its value would have been either $+1/2$ or $-1/2$," this is blocked by the single-framework rule applied to quantum counterfactuals (Chap. 19 of [8]). Thus the entire EPR "paradox" when analyzed from this point of view is nothing more than a particular application of the "paradox" that in Hilbert space quantum mechanics one cannot simultaneously assign values to S_z and S_x for a spin-half particle. The issue is entirely a matter of what one can say about the measurement of particle a. Particle b, together with entanglement, Bell inequalities, possible nonlocality, etc., are from this perspective entirely irrelevant. To be sure, entanglement, locality, and the like are in and of themselves interesting topics; for a detailed discussion from the consistent histories point of view, see [37,45] and Chaps. 23 and 24 of [8].

VI. CONCLUSION

We have shown that a satisfactory solution to the second measurement problem—inferring a prior microscopic state of affairs from the macroscopic outcome (pointer position) of a measurement described using quantum principles—exists for a significant class of projective and generalized (POVM) measurements. The approach using consistent histories is mathematically sound, gives reasonable physical results, and does not lead to paradoxes. Unlike current textbook treatments of measurements it makes no use of *ad hoc* principles and special rules that apply only when measurements are being made; instead the entire measurement process is analyzed using basic quantum principles that apply to all physical processes, whether microscopic or macroscopic.

A useful feature of the approach used here is the backwards map $Q^k = J^\dagger M^k J$, (35), relating a POVM element Q^k to the projector M^k on a subspace that corresponds to outcome ("pointer position") k. It is helpful for identifying an earlier microscopic property or properties that resulted in outcome k, even though it does not always give a precise answer. It is a significant addition to, while at the same time completely consistent with, the discussion of measurements in Chaps. 27 and 28 of [8]. And it would seem to be particularly useful for analyzing weak measurements in terms of physical properties rather than weak values, as illustrated by the simple example in Sec. V D.

The applications in Secs. V A to V D are relatively simple illustrations of the measurement formalism in Sec. IV, but the last two applications, Secs. V E and V F, address issues about which there is quite a bit of confusion in the published literature. Claims that quantum mechanics is "contextual" are incorrect if that term is interpreted in the sense introduced by Bell and used by Mermin. This has been pointed out previously [9], but one may hope that the quite specific example worked out in Sec. V E will result in a more precise definition of the term "contextual" on the part those who claim that quantum mechanics is contextual, or perhaps the withdrawal or modification of these claims. While the nonexistence of nonlocal influences in Bohm's version of the Einstein-Podolsky-Rosen paradox has been pointed out previously (see [37] and the references given there), the analysis in Sec. V F should help to further pin down the source of Bell's mistake: he did not have a solution to the second measurement problem (or, for that matter, the first; see [34]).

It is worth listing the fundamental quantum principles which make the consistent histories analysis possible. First, as we learned from von Neumann (Sec. II.5 of [16]), quantum properties (attributes of a physical system that can be true or false) correspond to subspaces of the quantum Hilbert space: no need for additional "hidden variables." Next, following a proposal by Born [46], quantum time dependence is inherently stochastic: Schrödinger's unitary time evolution should be used for calculating probabilities of events rather than determining them. Stochastic quantum time development can be described using histories represented by tensor products on a history Hilbert space, as first pointed out by Isham [47]. Assigning probabilities to quantum histories of a closed system using the extended Born rule requires the use of sample spaces satisfying consistency conditions; those used here are the medium decoherence conditions of Gell-Mann and Hartle [48].

Finally, a key principle that makes a clean break with classical thinking, and hence is often misunderstood by critics of consistent histories, is the abandonment of what elsewhere (Sec. 27.3 of [8]) has been called the *principle of unicity*: the idea that the universe, or at least that part of it which forms a closed physical system, must at any given time be in a single, well-defined physical state, a single point in a classical phase space. By contrast, the consistent histories approach gives the physicist liberty to construct alternative quantum descriptions—frameworks—which are incompatible with one another (and thus cannot be combined; the single-framework rule), each of which can make an equal claim to describing some aspect of physical reality. That freedom, discussed in greater detail in Ref. [15], is important for resolving both the first and the second measurement problem. As for the first problem, there is nothing fundamentally wrong with using unitary time evolution leading to a superposition state of different pointer positions, but this is of no use for discussing the measurement as having specific outcomes. Once unicity has been abandoned there is a perfectly good framework in which the pointer takes well-defined positions, each with some probability. As for the second problem, the textbook procedure that employs unitary evolution up to the time when the particle begins to interact with the apparatus is perfectly good quantum mechanics, but claiming that this is the *only* valid quantum description stands in the way of reaching the conclusion, using an appropriate framework, that the apparatus constructed by a competent experimenter actually did measure what it was designed to measure.

A proper understanding of what it is that quantum measurements measure should lead to a better physical understanding of the quantum world, and will, one hopes, someday replace the unsatisfactory discussion of quantum principles found in current textbooks. Students find introductory quantum theory hard to understand both because the mathematics is unfamiliar and because its connection with physical concepts seems obscure. They are not helped by the way "measurement" suddenly shows up in an almost magical way in textbook quantum mechanics. Somehow it doesn't look like good physics. And it isn't. Students deserve something better.

ACKNOWLEDGMENT

The author is grateful to the anonymous referees for their careful reading of the text.

[1] L. Vaidman, Past of a quantum particle, Phys. Rev. A **87**, 052104, (2013).

[2] R. B. Griffiths, Particle path through a nested Mach-Zehnder interferometer, Phys. Rev. A **94**, 032115 (2016).

[3] Z.-H. Li, M. Al-Amri, and M. Suhail Zubairy, Direct counterfactual transmission of a quantum state, Phys. Rev. A **92**, 052315 (2015).

[4] L. Vaidman, Comment on "Direct counterfactual transmission of a quantum state", Phys. Rev. A **93**, 066301 (2016).

[5] Z.-H. Li, M. Al-Amri, and M. Suhail Zubairy, Reply to Comment on "Direct counterfactual transmission of a quantum state", Phys. Rev. A **93**, 066302 (2016).

[6] N. Gisin, arXiv:1701.08300.

[7] S. Weinberg, *Lectures on Quantum Mechanics*, 2nd ed. (Cambridge University Press, Cambridge, 2015).

[8] R. B. Griffiths, *Consistent Quantum Theory* (Cambridge University Press, Cambridge, 2002).

[9] R. B. Griffiths, Hilbert space quantum mechanics is noncontextual, Stud. Hist. Phil. Mod. Phys. **44**, 174 (2013).

[10] M. Jammer, *The Philosophy of Quantum Mechanics* (Wiley, New York, 1974).

[11] G. Bacciagaluppi and A. Valentini, *Quantum Theory at the Crossroads* (Cambridge University Press, New York, 2009).

[12] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics*, Vol. III, *Quantum Mechanics* (Addison-Wesley, Reading, MA, 1965), Chap. 1.

[13] J. A. Wheeler, The "past" and the "delayed-choice" double-slit experiment, in A. R. Marlow, editor, *Mathematical Foundations of Quantum Theory* (Academic Press, New York, 1978), pp. 9–48.

[14] R. B. Griffiths, The consistent histories approach to quantum mechanics, in *Stanford Encyclopedia of Philosophy* (2014), http://plato.stanford.edu/entries/qm-consistent-histories

[15] R. B. Griffiths, The new quantum logic, Found. Phys. **44**, 610 (2014).

[16] J. von Neumann, *Mathematische Grundlagen der Quantenmechanik* (Springer-Verlag, Berlin, 1932) [English translation by R. T. Beyer, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, 1955)].

[17] G. Birkhoff and J. von Neumann, The logic of quantum mechanics, Ann. Math. **37**, 823 (1936).

[18] P. Mittelstaedt, Quantum logic, in D. Greenberger, K. Hentschel, and F. Weinert, editors, *Compendium of Quantum Physics* (Springer-Verlag, Berlin, 2009), pp. 601–604.

[19] R. B. Griffiths, A consistent quantum ontology, Stud. Hist. Phil. Mod. Phys. **44**, 93 (2013).

[20] P. W. Shor, Algorithms for quantum computation: Discrete logarithms and factoring, in edited by S. Goldwasser, *Proceedings of the 35th Annual Symposium on Foundations of Computer Science* (IEEE Computer Society Press, Los Alamitos, CA, 1994), pp. 124–134.

[21] E. Gerjuoy, Shor's factoring algorithm and modern cryptography: An illustration of the capabilities inherent in quantum computers, Am. J. Phys. **73**, 521 (2005).

[22] G. Lüders, Concerning the state-change due to the measurement process, Ann. Phys. (Leipzig) **15**, 663 (2006).

[23] P. Busch and P. Lahti, Lüder's rule, in D. Greenberger, K. Hentschel, and F. Weinert, editors, *Compendium of Quantum Physics* (Springer-Verlag, Berlin, 2009), pp. 356–358.

[24] B. Tamir and E. Cohen, Introduction to weak measurements and weak values, Quanta **2**, 7 (2013).

[25] B. E. Y. Svensson, Pedagogical review of quantum measurement theory with an emphasis on weak measurements, Quanta **2**, 18 (2013).

[26] Quantum tomography, in Wikipedia, en.wikipedia.org/wiki/Quantum_tomography

[27] A. W. Simmons, J. J. Wallman, H. Pashayan, S. D. Bartlett, and T. Rudolph, Contextuality under weak assumptions, New J. Phys. **19**, 033030 (2017).

[28] J. S. Bell, On the problem of hidden variables in quantum mechanics, Rev. Mod. Phys. **38**, 447 (1966) [reprinted in J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics*, 2nd ed. (Cambridge University Press, Cambridge, 2004), pp. 1–13].

[29] N. D. Mermin, Hidden variables and the two theorems of John Bell, Rev. Mod. Phys. **65**, 803 (1993).

[30] A. Peres, *Quantum Theory: Concepts and Methods* (Kluwer Academic Publishers, Dordrecht, The Netherlands, 1993).

[31] R. B. Griffiths, Consistent quantum counterfactuals, Phys. Rev. A **60**, R5(R) (1999).

[32] F. Laloë, Do we really understand quantum mechanics? Strange correlations, paradoxes, and theorems, Am. J. Phys. **69**, 655 (2001).

[33] F. Laloë, *Do We Really Understand Quantum Mechanics?* (Cambridge University Press, Cambridge, 2012).

[34] J. S. Bell, Against measurement, in A. I. Miller, editor, *Sixty-Two Years of Uncertainty* (Plenum Press, New York, 1990), pp. 17–31 [reprinted in J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics*, 2nd ed. (Cambridge University Press, Cambridge, 2004), pp. 213–231].

[35] S. Goldstein, Bohmian mechanics, in *Stanford Encyclopedia of Philosophy* (2012), http://plato.stanford.edu/entries/qm-bohm

[36] A. Fine, Joint distributions, quantum correlations, and commuting observables, J. Math. Phys. **23**, 1306 (1982).

[37] R. B. Griffiths, Quantum locality, Found. Phys. **41**, 705 (2011).

[38] R. W. Spekkens, Evidence for the epistemic view of quantum states: A toy theory, Phys. Rev. A **75**, 032110 (2007).

[39] M. S. Leifer, Is the quantum state real? A review of ψ-ontology theorems, Quanta **3**, 67 (2014).

[40] D. Jennings and M. Leifer, No return to classical reality, Contemp. Phys. **57**, 60 (2016).
[41] S. Kochen and E. P. Specker, The problem of hidden variables in quantum mechanics, J. Math. Mech. **17**, 59 (1967).
[42] S. Abramsky, R. S. Barbosa, and S. Mansfield, The contextual fraction as a measure of contextuality, Phys. Rev. Lett. **119**, 050504 (2017).
[43] A. Einstein, B. Podolsky, and N. Rosen, Can quantum-mechanical description of physical reality be considered complete? Phys. Rev. **47**, 777 (1935).
[44] D. Bohm, *Quantum Theory* (Prentice Hall, Englewood Cliffs, NJ, 1951), Chap. 22.
[45] R. B. Griffiths, EPR, Bell, and quantum locality, Am. J. Phys. **79**, 954 (2011).
[46] M. Born, Zur Quantenmechanik der Stoßvorgänge, Z. Phys. **37**, 863 (1926).
[47] C. J. Isham, Quantum logic and the histories approach to quantum theory, J. Math. Phys. **35**, 2157 (1994).
[48] M. Gell-Mann and J. B. Hartle, Classical equations for quantum systems, Phys. Rev. D **47**, 3345 (1993).